



# STIC EIC 2100 103820

## Search Request Form 57

Today's Date:

What date would you like to use to limit the search?

Priority Date:

Other:

Name TO, BAOQUOC

AU 2172 Examiner # 78889

Room # 4A42 Phone 305-1949

Serial # 89913960

Format for Search Results (Circle One):

☒ PAPER

☐ DISK

☐ EMAIL

Where have you searched so far?

☒ USP

☒ DWPI

☒ EPO

☒ JPO

☐ ACM

☒ IBM

☒ TDB

☐ IEEE

☐ INSPEC

☐ SPI

☐ Other

Is this a "Fast & Focused" Search Request? (Circle One) ☒ YES ☐ NO

A "Fast & Focused" Search is completed in 2-3 hours (maximum). The search must be on a very specific topic and meet certain criteria. The criteria are posted in EIC2100 and on the EIC2100 NPL Web Page at <http://ptoweb/patents/stic/stic-tc2100.htm>.

What is the topic, novelty, motivation, utility, or other specific details defining the desired focus of this search? Please include the concepts, synonyms, keywords, acronyms, definitions, strategies, and anything else that helps to describe the topic. Please attach a copy of the abstract, background, brief summary, pertinent claims and any citations of relevant art you have found.

Vector database and indexing; vector indexing  
calculating a declination  $(V_b \cdot P_d) / (|V_b| * |P_d|)$

Claim 1

09-12-03 P02:19 IN

STIC Searcher Geoffrey St. Leger

Phone 308-7800

Date picked up 9/22/3

Date Completed 9/22/3





# STIC Search Report

## EIC 2100

STIC Database Tracking Number: 103820

**TO: Baoquoc To**  
**Location: 4A42**  
**Art Unit : 2172**  
**Monday, September 22, 2003**

**Case Serial Number: 09/913960**

**From: Geoffrey St. Leger**  
**Location: EIC 2100**  
**PK2-4B30**  
**Phone: 308-7800**

**geoffrey.stleger@uspto.gov**

### Search Notes

Dear Examiner To,

Attached please find the results of your search request for application 09/913960. I searched Dialog's foreign patent files and technical databases; along with ACM and the Internet.

Please let me know if you have any questions.

Regards,

Geoffrey St. Leger  
4B30/308-7800



# STIC Search Results Feedback Form

**EIC 2100**

Questions about the scope or the results of the search? Contact *the EIC searcher or contact:*

Anne Hendrickson, EIC 2100 Team Leader  
308-7831, CPK2-4B40

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 3730

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

*Types of relevant prior art found:*

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

**Comments:**

Drop off or send completed forms to STIC/EIC2100 CPK2-4B40



File 348:EUROPEAN PATENTS 1978-2003/Sep W02

(c) 2003 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20030918,UT=20030911

(c) 2003 WIPO/Univentio

Set	Items	Description
S1	106770	VECTOR? ?
S2	1786	S1(5N)(INDEX??? OR INDICE? ?)
S3	3924	SIMILARIT???(5N)(SEARCH??? OR QUERIE? ? OR QUERY???)
S4	394	S1(5N)DECOMPOS?
S5	485135	IMAGE? ? OR PHOTO? ? OR PHOTOGRAPH? ? OR PICTURE? ?
S6	129549	DATABASE? ? OR DATA()BASE? ? OR REPOSITOR???
S7	14	S2(S)S3
S8	2548	S1(10N)(INDEX??? OR INDICE? ?)
S9	18	S8(S)S3
S10	9	S8(S)S3(S)S4:S6
S11	18	S9:S10

11/5,K/1 (Item 1 from file: 348)  
DIALOG(R)File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01429995

**Database indexing using a tree structure**

**Datenbankindizierung unter Benutzung einer Baumstruktur**

**Indexage d'une base de donnees en utilisant une structure en arbre**

**PATENT ASSIGNEE:**

SAMSUNG ELECTRONICS CO. LTD., (1093724), 416 Maetan-dong, Kwonsun-gu,  
Suwon-city, Kyungki-do, (KR), (Applicant designated States: all)  
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, (2137865), 12th Floor, 1111  
Franklin Street, Oakland, California 94607-5200, (US), (Applicant  
designated States: all)

**INVENTOR:**

Choi, Yang-lim, 102-1112 Wooman Sunkyung Apt., 105 Wooman-dong,,  
Paldal-gu, Suwon-city, Kyungki-do, (KR)  
Huh, Youngsik, (105) 1032-4 Youngtong-dong, Paldal-gu, Suwon-city,  
Kyungki-do, (KR)  
Manjunath, Bangalore S., Dept. Electr.& Comp.Eng., University of  
California, Santa Barbara, CA 93106-9560, (US)  
Chandrasekaran, Shiv, Dept. Electr.& Comp. Eng. University of California,  
Santa Barbara, CA 93106-9560, (US)

**LEGAL REPRESENTATIVE:**

Geary, Stuart Lloyd et al (79361), Venner, Shipley & Co., 20 Little  
Britain, London EC1A 7DH, (GB)

PATENT (CC, No, Kind, Date): EP 1207464 A2 020522 (Basic)

APPLICATION (CC, No, Date): EP 2001301675 010223;

PRIORITY (CC, No, Date): US 252391 P 001115; KR 2079180 001220

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI;  
LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

**ABSTRACT EP 1207464 A2**

A database of data items, such as images or video, is indexed by using  
a tree structure. Each node (302, 304, 306, 308, 310, 312, 314, 316, 318,  
320) of the tree structure relates to a region (202, 204, 206, 208,  
2061)) + 2062)) + 2063)),  
2064(underscore))2+2064(underscore)3))+2064(underscore)4)),  
2064(underscore)1)), 2081)), 2082))+2083))+2084))) in a feature vector  
region. At least one of the terminal nodes (304, 312, 314, 316, 318, 320)  
of the tree indexes a composite region (2061))+2062))+2063)),  
2064(underscore)2))+2064(underscore)3))+2064(underscore)4)),  
2082))+2083))+2084))) formed by combining a plurality of low population  
regions from the same index level.

ABSTRACT WORD COUNT: 93

**NOTE:**

Figure number on first page: NONE

**LEGAL STATUS (Type, Pub Date, Kind, Text):**

Application: 020522 A2 Published application without search report  
LANGUAGE (Publication,Procedural,Application): English; English; English  
FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200221	1580
SPEC A	(English)	200221	2492
Total word count - document A			4072
Total word count - document B			0
Total word count - documents A + B			4072

...SPECIFICATION Furthermore, quicker retrievals can be provided by  
repeatedly updating the predetermined threshold value e used in the  
similarity measurement.

That is, a high-dimensional feature **vector** space **indexed** by the  
method of **indexing** a feature **vector** space according to the present  
invention can support functions such as **similarity search**, retrieval  
or browsing in a salable and efficient manner. Thus, even if the size of

a **database** increases, the time required for **similarity search** and retrieval does not increase as much.

Furthermore, the method of indexing and searching a feature vector space according to the present invention can be...

11/5,K/2 (Item 2 from file: 348)  
DIALOG(R)File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01429170

**Adaptive search method in feature vector space**

**Angepasste Suchmethode im Eigenschaftens-Vektorraum**

**Procede de recherche adaptive dans un espace de vecteurs de caracteristiques**

PATENT ASSIGNEE:

SAMSUNG ELECTRONICS CO., LTD., (2171363), 416, Maetan-dong, Kwonsun-gu, Suweon-city, Kyungki-do, (KR), (Applicant designated States: all)  
The Regents of the University of California, (2289354), 12th Floor, 1111 Franklin Street, Oakland, CA 94607-5200, (US), (Applicant designated States: all)

INVENTOR:

Choi, Yang-lim, 210-1509 Imae Hanshin Apt, 124 Imaechon, Imae-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, (KR)  
Huh, Youngsik, (105) 1032-4 Youngtong-dong, Paldal-gu, Suwon-city, Kyungki-do, (KR)  
Manjunath, Bangalore S., Dpt.of Electrical and, Computer Engineering, University of California, Santa Barbara, CA 93106-9560, (US)  
Wu, Peng, Department of Electrical and Computer, Engineering, University of California, Santa Barbara, CA 93106-9560, (US)

LEGAL REPRESENTATIVE:

Geary, Stuart Lloyd et al (79361), Venner, Shipley & Co., 20 Little Britain, London EC1A 7DH, (GB)

PATENT (CC, No, Kind, Date): EP 1205856 A2 020515 (Basic)

APPLICATION (CC, No, Date): EP 2001300749 010129;

PRIORITY (CC, No, Date): US 248012 P 001114; KR 2079181 001220

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

ABSTRACT EP 1205856 A2

An adaptive search method in feature **vector** space which can quickly search the feature **vector** space **indexed** based on approximation for a feature **vector** having features similar to a query vector according to a varying distance measurement is provided. The adaptive search method includes the steps of (a) performing a **similarity** measurement on a given **query** vector within the feature vector space, and (b) applying search conditions limited by the result of the similar measurement obtained in the step (a) and performing a changed **similarity** measurement on the given **query** vector. According to the adaptive search method, the number of candidate approximation regions is reduced during a varying distance measurement such as an on-line retrieval, which improves the search speed.

ABSTRACT WORD COUNT: 121

NOTE:

Figure number on first page: NONE

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 020515 A2 Published application without search report  
LANGUAGE (Publication,Procedural,Application): English; English; English  
FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200220	536
SPEC A	(English)	200220	1889
Total word count - document A			2425
Total word count - document B			0
Total word count - documents A + B			2425

...ABSTRACT A2

An adaptive search method in feature **vector** space which can quickly search the feature **vector** space **indexed** based on approximation for a feature **vector** having features similar to a query vector according to a varying distance measurement is provided. The adaptive search method includes the steps of (a) performing a **similarity** measurement on a given **query** vector within the feature vector space, and (b) applying search conditions limited by the result of the similar measurement obtained in the step (a) and performing a changed **similarity** measurement on the given **query** vector. According to the adaptive search method, the number of candidate approximation regions is reduced during a varying distance measurement such as an on-line...

11/5,K/3 (Item 3 from file: 348)  
DIALOG(R)File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01411371

Image object ranking

Objektordnung in Bildern

Rangement des objets dans des images

PATENT ASSIGNEE:

Texas Instruments Incorporated, (279078), 7839 Churchill Way, Mail  
Station 3999, Dallas, Texas 75251, (US), (Applicant designated States:  
all)

INVENTOR:

Ronk, Lawrence J., 6261 Shady Brook Ln.Apt.No.188, Dallas Texas 75206,  
(US)

Flinchbaugh, Bruce E., 6402 Riverview Lane, Dallas Texas 75248, (US)

LEGAL REPRESENTATIVE:

Holt, Michael (50422), Texas Instruments Ltd., EPD MS/13, 800 Pavilion  
Drive, Northampton Business Park, Northampton NN4 7YL, (GB)

PATENT (CC, No, Kind, Date): EP 1193648 A1 020403 (Basic)

APPLICATION (CC, No, Date): EP 2001000264 010702;

PRIORITY (CC, No, Date): US 215248 P 000630

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI;  
LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06T-007/20

ABSTRACT EP 1193648 A1

Automatic vision system object indexing and image database query system using both path-dependent and path-independent features of moving objects within a sequence of images. Feature vectors of both average over frames of an object traversing the field of view plus average over blocks of a grid for a path association. Color histograms may be an included feature.

ABSTRACT WORD COUNT: 58

NOTE:

Figure number on first page: 1

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 020403 A1 Published application with search report

Examination: 021204 A1 Date of request for examination: 20021004

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200214	283
SPEC A	(English)	200214	6925
Total word count - document A			7208
Total word count - document B			0
Total word count - documents A + B			7208

...SPECIFICATION the 10% mark are listed in table 5.

Modifications

The preferred embodiments may be modified in various ways while

retaining the aspects of video object **indexing** with feature **vectors** plus grid block sequences reflecting objects' paths of traversing the field of view, and the query method of feature vector and grid block sequence **similarity searching** metrics (including color histogram) for finding objects of interest.

For example, the path-independent and the path-dependent features could be varied; the averaging of...

11/5,K/4 (Item 4 from file: 348)  
DIALOG(R) File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01361784

**Method of indexing and similarity search in a feature vector space**  
**Methode zur Indizierung und Ähnlichkeitssuche in einem Feature-Vektor-Raum**  
**Procede d'indexation et recherche de similarite dans un espace de vecteurs de caracteristiques**

PATENT ASSIGNEE:

SAMSUNG ELECTRONICS CO. LTD., (1093724), 416 Maetan-dong, Kwonsun-gu, Suwon-city, Kyungki-do, (KR), (Applicant designated States: all)  
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, (2137865), 12th Floor, 1111 Franklin Street, Oakland, California 94607-5200, (US), (Applicant designated States: all)

INVENTOR:

Shin, Hyun-doo, 510-1302Mugigae Maeul Cheongku Apt, Kumi-dong, Pundang-gu, Seongnam-city, Kyungki-do, (KR)  
Choi, Yang-lim, 210-1509 Imae Hanshin Apt., 124 Imae-chon, Imae-dong Bundang-gu, Seongnam si Gyeonggi-do, (KR)  
Wu, Peng, Dpt.of Electr. and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, (US)  
Manjunath, Bangalore S., Dpt. of Electrical and, Computer Engineering, University of California, Santa Barbara, CA 93106-9560, (US)

LEGAL REPRESENTATIVE:

Geary, Stuart Lloyd et al (79361), Venner, Shipley & Co., 20 Little Britain, London EC1A 7DH, (GB)

PATENT (CC, No, Kind, Date): EP 1160690 A1 011205 (Basic)

APPLICATION (CC, No, Date): EP 2001302946 010329;

PRIORITY (CC, No, Date): US 208086 P 000531; KR 2048323 000821

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

ABSTRACT EP 1160690 A1

In the field of indexing multidimensional data, there has not been a satisfactory data structure to support the nearest neighbor (NN)-search efficiently when the feature **vectors** are not uniformly distributed.

A method of **indexing** data elements in a feature **vector** space comprises assigning data elements to first level **index** terms with a first course granularity and identifying first level index terms for which there are concentrations of elements and extending these first level index terms to provide a finer grained index term for the elements making up the concentrations of elements.

Additionally a method of **searching** for **similarity** in a feature vector data space is provided.

ABSTRACT WORD COUNT: 109

NOTE:

Figure number on first page: 1

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 011205 A1 Published application with search report

Examination: 020424 A1 Date of request for examination: 20020211

Examination: 020612 A1 Date of dispatch of the first examination report: 20020426

Change: 020717 A1 Inventor information changed: 20020529

LANGUAGE (Publication,Procedural,Application): English; English; English



#### FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200149	715
SPEC A	(English)	200149	1577
Total word count - document A			2292
Total word count - document B			0
Total word count - documents A + B			2292

#### Method of indexing and similarity search in a feature vector space

...ABSTRACT the field of indexing multidimensional data, there has not been a satisfactory data structure to support the nearest neighbor (NN)-search efficiently when the feature **vectors** are not uniformly distributed.

A method of **indexing** data elements in a feature **vector** space comprises assigning data elements to first level **index** terms with a first course granularity and identifying first level index terms for which there are concentrations of elements and extending these first level index terms to provide a finer grained index term for the elements making up the concentrations of elements.

Additionally a method of **searching** for **similarity** in a feature vector data space is provided.

...SPECIFICATION distribution of feature vector data in a feature vector data space to cope with the concentration of feature vector data.

A method of performing a **similarity search** on a feature **vector** data space which has been hierarchically **indexed** according to the **indexing** method of a feature **vector** space described with reference to Figure 1, will now be described.

Feature vectors in each cell on which feature vectors are concentrated in the feature...

...CLAIMS in each cell, on which feature vectors are concentrated, using the vector approximation file and a corresponding sub-vector approximation file.

15. A method of **searching** for **similarity** in a feature **vector** data space in which feature **vectors** are **indexed**, the method comprising the step of (a) performing a **similarity search** in the feature **vector** data space which has been **indexed** by determining whether cells on which feature **vectors** are concentrated exist and hierarchically **indexing** feature **vectors** in the cells on which it is determined that feature **vectors** are concentrated according to a predetermined **indexing** method.

16. The method of claim 15, wherein the step (a) is performed based on a nearest neighbor search.

11/5,K/9 (Item 4 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2003 WIPO/Univentio. All rts. reserv.

00924616

#### LEUKOCYTE EXPRESSION PROFILING

#### EVALUATION DU NIVEAU D'EXPRESSION LEUCOCYTAIRE

Patent Applicant/Assignee:

EXPRESSION DIAGNOSTICS INC, 384 Oyster Point Boulevard, Suite No. 6,  
South San Francisco, CA 94080, US, US (Residence), US (Nationality),  
(For all designated states except: US)

Patent Applicant/Inventor:

WOHLGEMUTH Jay, 664 Hamilton Avenue, Palo Alto, CA 94301, US, US  
(Residence), US (Nationality), (Designated only for: US)

FRY Kirk, 2604 Ross Road, Palo Alto, CA 94303, US, US (Residence), US  
(Nationality), (Designated only for: US)

MATCUK George, 141C Escondido Village, Stanford, CA 94305, US, US  
(Residence), US (Nationality), (Designated only for: US)

ALTMAN Peter, 717 Evelyn Avenue, Albany, CA 94706, US, US (Residence), US  
(Nationality), (Designated only for: US)

PRENTICE James, 120 Dolores Street, San Francisco, CA 94103, US, US  
(Residence), US (Nationality), (Designated only for: US)  
PHILLIPS Julie, 1090 Mirador Terrace, Pacifica, CA 94044, US, US  
(Residence), US (Nationality), (Designated only for: US)  
LY Ngoc, 2000 Crystal Springs Road 15-14, San Bruno, CA 94066, US, US  
(Residence), US (Nationality), (Designated only for: US)  
WOODWARD Robert, 1828 Rheem Court, Pleasanton, CA 94588, US, US  
(Residence), US (Nationality), (Designated only for: US)  
QUINTERMOUS Thomas, 44 El Rey Road, Portola Valley, CA 94028, US, US  
(Residence), US (Nationality), (Designated only for: US)  
JOHNSON Frances, 44 El Rey Road, Portola Valley, CA 94028, US, US  
(Residence), US (Nationality), (Designated only for: US)

Legal Representative:

WARD Michael R (et al) (agent), Morrison & Foerster LLP, 425 Market  
Street, San Francisco, CA 94105-2482, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200257414 A2-A3 20020725 (WO 0257414)

Application: WO 2001US47856 20011022 (PCT/WO US0147856)

Priority Application: US 2000241994 20001020; US 2001296764 20010608

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU

CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP

KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU

SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: C07H-021/04

International Patent Class: C12Q-001/68

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 314482

English Abstract

Leukocyte gene expression profiling is utilized to identify  
oligonucleotides from gene expression candidate libraries. The expression  
libraries are generally immobilized on an array. Diagnostic  
oligonucleotide sets for analysis of leukocyte-related diseases are  
described.

French Abstract

L'invention concerne l'evaluation du niveau d'expression genique d'un  
leucocyte utilise pour identifier des oligonucleotides a partir de  
bibliotheques candidates d'expression genique. Ces bibliotheques  
d'expression sont generalement immobilisees sur une matrice. L'invention  
concerne egalement un oligonucleotide de diagnostic regle de facon a  
analyser des maladies associees a un leucocyte.

Legal Status (Type, Date, Text)

Publication 20020725 A2 Without international search report and to be  
republished upon receipt of that report.

Search Rpt 20020926 Late publication of international search report

Republication 20020926 A3 With international search report.

Republication 20020926 A3 Before the expiration of the time limit for  
amending the claims and to be republished in the  
event of the receipt of amendments.

Search Rpt 20020926 Late publication of international search report

Examination 20030213 Request for preliminary examination prior to end of  
19th month from priority date

Correction 20030912 Corrected version of Pamphlet: pages 1/10-10/10,  
drawings, replaced by new pages 1/11-11/11; due to  
late transmittal by the receiving Office

Republication 20030912 A3 With international search report.

Fulltext Availability:

## Detailed Description

### Detailed Description

... A SOLID

SUBSTRATE FOR USE IN NUCLEIC ACID HYBRIDIZATION ASSAYS" to Bahl et al., issued June 1 @ 1993; US Patent No. 5,707,807 "MOLECULAR

#### INDEXING

FOR EXPRESSED GENE ANALYSIS" to Kato, issued January 13, 1998; US Patent No. 5,807,522 "METHODS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES" to Brown...multisequence file with the appropriate labels for each clone in the headers for subsequent automated analysis.

104

Initially, known sequences were analyzed by pair wise **similarity searching** using the blastn option of the blastall program obtained from the National Center for Biological Information, National Library of Medicine, National Institutes of Health (NCBI...

...were removed from the sequences.

Messenger RNA contains repetitive elements that are found in genomic DNA.

These repetitive elements lead to false positive results in **similarity searches** of **query** mRNA sequences versus known mRNA and EST databases. Additionally, regions of low information content (long runs of the same nucleotide, for example) also result in...design a probe for expression analysis and further approaches are taken to identify the gene or predicted gene that corresponds to the cDNA sequence, including **similarity searches** of other databases, molecular cloning, and Rapid Amplification of cDNA Ends (RACE).

In some cases, the process of analyzing many unknown sequences with BLASTN was...the peptide-predicting algorithms used to create the two sequences, but the homology between the two is significant.

BLASTP and TBLASTN were also used to **search** for sequence **similarities** in the SWISS-PROT, TrEMBL, GenBank Translated, and PDB databases. Matches to several proteins were found, among them a tumor cell suppression protein, HTS 1...

...used to conduct farther domain and motif analysis. The Prosite search generated many recognized protein domains. A BLASTP search was performed to identify areas of **similarity** between the protein **query** sequence and PRINTS, a protein database of protein fingerprints, groups of motifs that together form a characteristic signature of a protein family. In this case...PDB databases. No significant matches were found in any of these, so no gene identity or tertiary structure was discovered.

The peptide sequence was also **searched** for **similarity** to known domains and motifs using BLASTP with the Prosite, Blocks, Pfam, and ProDorn databases. The searches produced no significant alignments to known domains. BLASTP...

11/5,K/18 (Item 13 from file: 349)  
DIALOG(R)File 349:PCT FULLTEXT  
(c) 2003 WIPO/Univentio. All rts. reserv.

00328285 \*\*Image available\*\*

#### A DATABASE ACCESSING SYSTEM

#### SYSTEME D'ACCES A UNE BASE DE DONNEES

Patent Applicant/Assignee:

HELFGOTT & KARAS P C,  
PHONETIC SYSTEMS LTD,

Inventor(s):

GILAI Atzmon,  
RESNEKOV Hezi,

Patent and Priority Information (Country, Number, Date):

Patent: WO 9610795 A1 19960411

Application: WO 95US12600 19951002 (PCT/WO US9512600)

Priority Application: US 94317040 19941003

Designated States: AM AT AU BB BG BR BY CA CH CN CZ DE DK EE ES FI GB GE HU  
IS JP KE KG KP KR KZ LK LR LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU  
SD SE SG SI SK TJ TM TT UA UG UZ VN KE MW SD SZ UG AT BE CH DE DK ES FR  
GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN ML MR NE SN TD TG

Main International Patent Class: G06F-017/30

International Patent Class: G06F-17:20; G06F-17:21

Publication Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 9993

English Abstract

A database accessing system for processing a request to access a database including a multiplicity of entries, each entry including at least one word, the request including a sequence of representations of possibly erroneous user inputs (10), the system including a similar word finder operative (30), for at least one interpretation of each representation, to find at least one database word which is similar to that interpretation, and a database entry evaluator operative (50), for each database word found by the similar word finder, to assign similarity values for relevant entries in the database, the values representing the degree of similarity between each database entry and the request.

French Abstract

L'invention concerne un systeme d'accès a une base de donnees, permettant de traiter une demande d'accès a une base de donnees comprenant une multiplicité d'entrees, chaque entree comprenant au moins un mot, la demande comprenant une sequence de representations de donnees (10) d'utilisateur pouvant etre erronees. Le systeme comprend un identificateur (30) de mots similaires, capable d'effectuer au moins une interpretation de chaque representation, de maniere a trouver au moins un mot de la base de donnees qui est similaire a cette interpretation, et un evaluateur (50) des entrees a la base de donnees pour chaque mot de la base de donnees trouve par l'identificateur de mots similaires, permettant d'attribuer des valeurs de similarite a des entrees pertinentes dans la base de donnees, les valeurs representant le degre de similarite entre chaque entree de la base de donnees et la demande.

Fulltext Availability:

Detailed Description

Detailed Description

... complete original alphabet.

Each layer of the similarity search index contains the same dictionary words but in a different reduced alphabet.

Each word in the **similarity search index** is represented in **vector** format with a reduced alphabet, A lines. Fig. 13 illustrates an example of a "grapheme based" similarity search index for an English language dictionary. It...

File 347:JAPIO Oct 1976-2003/May(Updated 030902)

(c) 2003 JPO & JAPIO

File 350:Derwent WPIX 1963-2003/UD,UM &UP=200360

(c) 2003 Thomson Derwent

Set	Items	Description
S1	93114	VECTOR? ?
S2	582	S1(5N)(INDEX??? OR INDICE? ?)
S3	509	SIMILARIT???(5N)(SEARCH??? OR QUERIE? ? OR QUERY???)
S4	159	S1(5N)DECOMPOS?
S5	1539621	IMAGE? ? OR PHOTO? ? OR PHOTOGRAPH? ? OR PICTURE? ?
S6	119636	DATABASE? ? OR DATA()BASE? ? OR REPOSITOR???
S7	5	S2 AND S3
S8	4	S7 AND S4:S6
S9	793	S1(10N)(INDEX??? OR INDICE? ?)
S10	1752	S1 AND (INDEX??? OR INDICE? ?)
S11	8	S10 AND S3
S12	6	S11 AND S4:S6
S13	8	S11:S12

13/5/1 (Item 1 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2003 JPO & JAPIO. All rts. reserv.

07294818 \*\*Image available\*\*  
ADAPTIVE RETRIEVING METHOD IN CHARACTERISTIC VECTOR SPACE

PUB. NO.: 2002-163292 [JP 2002163292 A]  
PUBLISHED: June 07, 2002 (20020607)  
INVENTOR(s): SAI RYORIN  
HUH YOUNGSIK  
MANJUNATH BANGALORE S  
WU PENG  
APPLICANT(s): SAMSUNG ELECTRONICS CO LTD  
REGENTS OF THE UNIV OF CALIFORNIA  
APPL. NO.: 2001-220128 [JP 20011220128]  
FILED: July 19, 2001 (20010719)  
PRIORITY: 00 248012 [US 2000248012], US (United States of America),  
November 14, 2000 (20001114)  
00 200079181 [KR 200079181], KR (Korea) Republic of, December  
20, 2000 (20001220)  
INTL CLASS: G06F-017/30; G06F-017/16; G06F-017/17

#### ABSTRACT

PROBLEM TO BE SOLVED: To provide a method of quickly retrieving the characteristic **vector** having the characteristic similar to the query **vector** by measuring a distance variable in a **vector** space **indexed** on the basis of the approximation.

SOLUTION: This adaptive retrieving method includes (a) a stage for measuring the similarity in the characteristic **vector** space with respect to the given query **vector**, and (b) a stage for measuring the changed **similarity** with respect to the given **query vector** by applying a retrieving condition limited by a result of the measurement of the similarity obtained in the (a) stage. As the number of the candidate approximate areas is small in measuring the variable distance such as on-line retrieving, a retrieving speed can be improved.

COPYRIGHT: (C)2002,JPO

13/5/2 (Item 2 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2003 JPO & JAPIO. All rts. reserv.

06736603 \*\*Image available\*\*  
**SIMILARITY SEARCHING** METHOD FOR VIDEO, METHOD FOR PRESENTING VIDEO  
WITHIN VIDEO BROWSER, METHOD FOR PRESENTING VIDEO WITHIN INTERFACE OF WEB  
BASE, COMPUTER READABLE RECORDING MEDIUM AND COMPUTER SYSTEM

PUB. NO.: 2000-322450 [JP 2000322450 A]  
PUBLISHED: November 24, 2000 (20001124)  
INVENTOR(s): JONATHAN T FOOTE  
ANDREAS GARGENSHORN  
WILCOX LYNN D  
APPLICANT(s): FUJI XEROX CO LTD  
APPL. NO.: 2000-064979 [JP 200064979]  
FILED: March 09, 2000 (20000309)  
PRIORITY: 266558 [US 99266558], US (United States of America), March  
11, 1999 (19990311)  
INTL CLASS: G06F-017/30; G06T-007/00; G10L-015/14; G10L-015/00;  
H04N-005/76; H04N-005/91; H04N-007/30

#### ABSTRACT

PROBLEM TO BE SOLVED: To extremely accelerate the calculation of similarity and to enable the application of an interactive style by obtaining a reduced feature **vector** corresponding to each frame of a video and by calculating a similarity score while using this reduced feature **vector**

and statistical models.

SOLUTION: A video feature 208 is selected out of the matrix of transformation coefficients as a transformation coefficient at a coefficient position inside a transformation matrix shown as a video set for video classification. A classifier 206 receives the respective video features 208 and inputs these video features 208 to respective **image** class statistical models 202-205. As a result, the respective frames of a video file 201 are classified into **image** classes expressed by the **image** class statistical models 202-205. The corresponding **image** class determined so as to correspond to the frame of the video file 201 by the classifier 206 is **indexed** to a video 207 with a labeled class.

COPYRIGHT: (C)2000,JPO

13/5/3 (Item 1 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

015159878 \*\*Image available\*\*  
WPI Acc No: 2003-220406/200321  
XRPX Acc No: N03-175824

**Media objects classification system e.g. for digital audio files  
calculates query vector using index calculated after associating  
subsets of media objects into clusters**

Patent Assignee: COHEN M (COHE-I)

Inventor: COHEN M

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20020184193	A1	20021205	US 2001867774	A	20010530	200321 B

Priority Applications (No Type Date): US 2001867774 A 20010530

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20020184193	A1		6	G06F-007/00	

Abstract (Basic): US 20020184193 A1

NOVELTY - An electronic processor associates subsets of media objects into one or more clusters of dissimilar objects for calculating at least one **index** of one cluster. The **similarity** of the **query vector** is calculating using the **index**.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) method for constructing an **index** structure for **database** ;  
and

(2) method for searching a **database** for similar objects.

USE - For classification media objects e.g. digital audio files, electronic representation of audio visual works.

ADVANTAGE - Number similarity comparisons required to locate most similar **vector** in a combination is substantially reduced. Works well with **vectors** of very high dimensionality, thus solving dimensionality problems.

DESCRIPTION OF DRAWING(S) - The figure shows the graph explaining the relationship between clusters and threshold.

pp; 6 DwgNo 1/1

Title Terms: MEDIUM; OBJECT; CLASSIFY; SYSTEM; DIGITAL; AUDIO; FILE;

CALCULATE; QUERY; **VECTOR** ; **INDEX** ; CALCULATE; AFTER; ASSOCIATE; SUBSET;  
MEDIUM; OBJECT; CLUSTER

Derwent Class: T01

International Patent Class (Main): G06F-007/00

File Segment: EPI

13/5/5 (Item 3 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

014822069 \*\*Image available\*\*

WPI Acc No: 2002-642775/200269

XRPX Acc No: N02-508096

**Semantic content representing method for retrieving documents in computer system, involves performing singular value decomposition and dimensionality reduction of matrix to form latent semantic indexed vector space**

Patent Assignee: SCI APPL INT CORP (SCIT-N)

Inventor: BRADFORD R B; WNEK J

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20020103799	A1	20020801	US 2000251423	A	20001206	200269 B
			US 2001683263	A	20011205	

Priority Applications (No Type Date): US 2000251423 P 20001206; US 2001683263 A 20011205

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

US 20020103799	A1		15	G06F-007/00	Provisional application US 2000251423
----------------	----	--	----	-------------	---------------------------------------

Abstract (Basic): US 20020103799 A1

NOVELTY - A two dimensional matrix with columns representing documents and rows representing terms which include n-tuple term occurring in documents and elements related with number of occurrences of the term in the document, is formed. A latent semantic **indexed vector** space is formed by performing singular value decomposition and dimensionality reduction of the matrix.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) Conceptual **similarity** determining method for **query** and reference documents in computer system;

(2) Conceptual similarity determining method for subject and reference documents; and

(3) Query document representing method.

USE - For retrieving data stored in **database** or in computer files, for collecting material of conceptual relatedness for proposal preparation, research management, legal brief development, document declassification.

ADVANTAGE - The method searches a collection of one or more documents efficiently based on conceptual content even at sub-document level. The two documents which closely related by the concept is identified easily by the usage of different synonyms for documents in a **database**.

DESCRIPTION OF DRAWING(S) - The figure illustrates the singular value decomposition operation of the latent semantic content representing method.

pp; 15 DwgNo 2/4

Title Terms: CONTENT; REPRESENT; METHOD; RETRIEVAL; DOCUMENT; COMPUTER; SYSTEM; PERFORMANCE; SINGULAR; VALUE; DECOMPOSE; DIMENSION; REDUCE; MATRIX; FORM; LATENT; **INDEX** ; **VECTOR** ; SPACE

Derwent Class: T01

International Patent Class (Main): G06F-007/00

File Segment: EPI

13/5/6 (Item 4 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

014699551 \*\*Image available\*\*

WPI Acc No: 2002-520255/200255

XRPX Acc No: N02-411773

**Relevancy ranking and clustering system for e.g. Internet document queries involves conversion of queries into ontology-based predicts and**



**comparison against parsed documents**

Patent Assignee: CAUDILL M (CAUD-I); TSENG J C (TSEN-I); WANG L (WANG-I);  
SCI APPL INT CORP (SCIT-N)

Inventor: CAUDILL M; TSENG J C; WANG L

Number of Countries: 100 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200257961	A2	20020725	WO 2002US402	A	20020110	200255 B
US 20020129015	A1	20020912	US 2001761188	A	20010118	200262

Priority Applications (No Type Date): US 2001761188 A 20010118

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

WO 200257961	A2	E	89	G06F-017/30	
--------------	----	---	----	-------------	--

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA  
CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN  
IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ  
OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG UZ VN YU ZA  
ZM ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR  
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW

US 20020129015	A1			G06F-007/00	
----------------	----	--	--	-------------	--

Abstract (Basic): WO 200257961 A2

NOVELTY - Relevance of document to user's **query** is determined by calculating **similarity** coefficient based on structures of pair of query predicates and document predicates. Documents are autonomously clustered (140) using self-organizing neural network that provides coordinate system that makes judgments in non-subjective fashion.

DETAILED DESCRIPTION - System determines relevance of document relative to user's query using comparison process. Input queries are parsed into query predicate structures using an ontological parser. The ontological parser parses a set of known documents to generate document predicate structures. A comparison of each query predicate structure with each document predicate structure is performed to determine a matching degree, represented by a real number. A multilevel modifier strategy is implemented to assign different relevance values to the different parts of each predicate structure match to calculate the predicate structure's matching degree.

INDEPENDENT CLAIMS are also included for the following:

- (1) a clustering method using parsing and vectorizing,
- (2) a method of vectorizing a set of document predicate structures,
- (3) a relevancy ranking system,
- (4) a relevancy ranking system with a neural network,
- (5) a clustering system,
- (6) a question and answering system.

USE - Relevancy ranking and clustering system for document queries, **indexing** and retrieval including on the Internet.

ADVANTAGE - The system automates a document query process and enables the user to provide feedback in order to fine-tune the search process. The number **vectors** used for text representation are ontologically generated concept representations, with meaningful numerical relationships so closely related concepts have numerically similar representations while independent concepts have numerically different representations. Also, the concepts represented are in numerical form as part of complete predicate structures, rather than simple independent words. The vectorization method provides a way to represent both long and short queries with **vector** representations with same dimensions that permits faster clustering. The method also permits comparisons of large-scale patterns across the whole document rather than moving between small windows.

DESCRIPTION OF DRAWING(S) - The block diagram represents a relevancy ranking system.

Document clustering (140)

pp; 89 DwgNo 1/16

Title Terms: RANK; SYSTEM; DOCUMENT; QUERY; CONVERT; QUERY; BASED; PREDICT;  
COMPARE; DOCUMENT

Derwent Class: T01

International Patent Class (Main): G06F-007/00; G06F-017/30  
File Segment: EPI

13/5/7 (Item 5 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

014111885 \*\*Image available\*\*  
WPI Acc No: 2001-596097/200167  
XRPX Acc No: N01-444347

Vector index creating and similar vector searching method from  
vector database according to inner product criteria  
Patent Assignee: MATSUSHITA ELECTRIC IND CO LTD (MATU ); KANNO Y (KANN-I)  
Inventor: KANNO Y  
Number of Countries: 095 Number of Patents: 005  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200146858	A1	20010628	WO 2000JP9079	A	20001221	200167 B
AU 200123993	A	20010703	AU 200123993	A	20001221	200167
EP 1204032	A1	20020508	EP 2000987677	A	20001221	200238
			WO 2000JP9079	A	20001221	
US 20020178158	A1	20021128	WO 2000JP9079	A	20001221	200281
			US 2001913960	A	20010821	
JP 2001547305	X	20030610	WO 2000JP9079	A	20001221	200339
			JP 2001547305	A	20001221	

Priority Applications (No Type Date): JP 99363058 A 19991221  
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200146858	A1	J	96	G06F-017/30	
Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW					
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW					
AU 200123993	A			G06F-017/30	Based on patent WO 200146858
EP 1204032	A1	E		G06F-017/30	Based on patent WO 200146858
Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR					
US 20020178158	A1			G06F-007/00	
JP 2001547305	X			G06F-017/30	Based on patent WO 200146858

Abstract (Basic): WO 200146858 A1

NOVELTY - Similar **vectors** are fast retrieved from a **vector database** of several hundreds of dimensions with reference to a single **vector index** according to the criterion of the inner product or distance, after specifying the **similarity search** range and the maximum number of similar **vectors** to be retrieved. For creating the **vector index**.

USE - **Vector index** creating and similar **vector** searching method from **vector database** according to inner product criteria

DESCRIPTION OF DRAWING(S) - **Vector database** (101)

Sub- **vector** calculating means (102)

Norm distribution determining means (103)

Norm section list (104)

Area number calculating means (105)

Argument distribution determining means (106)

Argument section list (107)

Norm section number calculating means (108)

Argument section number calculating means (109)

**Index** data calculating means (110)

**Index** creating means (111)

**Vector index** (112)

pp; 96 DwgNo 1/29

Title Terms: **VECTOR** ; **INDEX** ; **SIMILAR** ; **VECTOR** ; **SEARCH** ; **METHOD** ; **VECTOR**  
; **DATABASE** ; **ACCORD** ; **INNER** ; **PRODUCT** ; **CRITERIA**

Derwent Class: T01  
International Patent Class (Main): G06F-007/00; G06F-017/30  
File Segment: EPI

13/5/8 (Item 6 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

011385393 \*\*Image available\*\*  
WPI Acc No: 1997-363300/199733  
XRPX Acc No: N97-302109

Method of indexing database of stored object - involves conducting  
similarity search on indexed set of truncated transformed feature  
vectors to retrieve set of vectors which represent super-set of  
objects including desired objects and false positives

Patent Assignee: INT BUSINESS MACHINES CORP (IBMC )  
Inventor: AGRAWAL R; EQUITZ W R; FALOUTSOS C; FLICKNER M D; SWAMI A N  
Number of Countries: 001 Number of Patents: 001  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 5647058	A	19970708	US 9366759	A	19930524	199733 B
			US 96607922	A	19960228	

Priority Applications (No Type Date): US 9366759 A 19930524; US 96607922 A  
19960228

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 5647058	A		13		Cont of application US 9366759

Abstract (Basic): US 5647058 A

The method involves applying a set of feature extraction functions to extract a set of feature **vectors** from the stored objects in the **database**. The set of feature extraction functions has a similarity measure applicable to the stored objects. The set of extracted feature **vectors** are transformed using an orthonormal transform such that the similarity measure is preserved.

The transformed feature **vectors** are truncated such that the entries which contribute little to the information of the transformed **vectors** are removed. The truncated feature **vectors** are **indexed** using a non-sequential point-access-method (PAM). A **similarity search** on the **indexed** set of truncated transformed feature **vectors** is conducted to retrieve a set of **vectors** which represent a superset of objects including desired objects and false positives. A secondary search is performed on the retrieved set of **vectors** to eliminate the false positives.

ADVANTAGE - Achieves efficient and complete retrieval from **database** of high dimensionality points while guaranteeing completeness and reduces prosperity for false positives.

Dwg.3/3

Title Terms: METHOD; **INDEX** ; **DATABASE** ; STORAGE; OBJECT; CONDUCTING;  
SIMILAR; SEARCH; **INDEX** ; SET; TRUNCATE; TRANSFORM; FEATURE; **VECTOR** ;  
RETRIEVAL; SET; **VECTOR** ; REPRESENT; SUPER; SET; OBJECT; OBJECT; FALSE;  
POSITIVE

Derwent Class: T01  
International Patent Class (Main): G06F-017/30  
File Segment: EPI

File 347:JAPIO Oct 1976-2003/May(Updated 030902)  
(c) 2003 JPO & JAPIO  
File 350:Derwent WPIX 1963-2003/UD,UM &UP=200360  
(c) 2003 Thomson Derwent  
File 348:EUROPEAN PATENTS 1978-2003/Sep W02  
(c) 2003 European Patent Office  
File 349:PCT FULLTEXT 1979-2002/UB=20030918,UT=20030911  
(c) 2003 WIPO/Univentio

Set	Items	Description
S1	312	AU=KANNO Y?
S2	4	S1 AND VECTOR? ?(5N) (INDEX??? OR INDICE? ?)

2/5/1 (Item 1 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2003 Thomson Derwent. All rts. reserv.

014111885 \*\*Image available\*\*

WPI Acc No: 2001-596097/200167

XRPX Acc No: N01-444347

Vector index creating and similar vector searching method from  
vector database according to inner product criteria

Patent Assignee: MATSUSHITA ELECTRIC IND CO LTD (MATU ); KANNO Y (KANN-I)

Inventor: KANNO Y

Number of Countries: 095 Number of Patents: 005

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200146858	A1	20010628	WO 2000JP9079	A	20001221	200167 B
AU 200123993	A	20010703	AU 200123993	A	20001221	200167
EP 1204032	A1	20020508	EP 2000987677	A	20001221	200238
			WO 2000JP9079	A	20001221	
US 20020178158	A1	20021128	WO 2000JP9079	A	20001221	200281
			US 2001913960	A	20010821	
JP 2001547305	X	20030610	WO 2000JP9079	A	20001221	200339
			JP 2001547305	A	20001221	

Priority Applications (No Type Date): JP 99363058 A 19991221

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 200146858 A1 J 96 G06F-017/30

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA  
CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP  
KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT  
RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR  
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200123993 A G06F-017/30 Based on patent WO 200146858

EP 1204032 A1 E G06F-017/30 Based on patent WO 200146858

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT  
LI LT LU LV MC MK NL PT RO SE SI TR

US 20020178158 A1 G06F-007/00

JP 2001547305 X G06F-017/30 Based on patent WO 200146858

Abstract (Basic): WO 200146858 A1

NOVELTY - Similar vectors are fast retrieved from a vector database  
of several hundreds of dimensions with reference to a single **vector**  
**index** according to the criterion of the inner product or distance,  
after specifying the similarity search range and the maximum number of  
similar vectors to be retrieved. For creating the **vector index**.

USE - **Vector index** creating and similar **vector** searching  
method from vector database according to inner product criteria

DESCRIPTION OF DRAWING(S) - Vector database (101)

Sub-vector calculating means (102)

Norm distribution determining means (103)

Norm section list (104)

Area number calculating means (105)

Argument distribution determining means (106)

Argument section list (107)

Norm section number calculating means (108)

Argument section number calculating means (109)

Index data calculating means (110)

Index creating means (111)

**Vector index** (112)

pp; 96 DwgNo 1/29

Title Terms: VECTOR; INDEX; SIMILAR; VECTOR; SEARCH; METHOD; VECTOR;

DATABASE; ACCORD; INNER; PRODUCT; CRITERIA

Derwent Class: T01

International Patent Class (Main): G06F-007/00; G06F-017/30

File Segment: EPI

2/5/2 (Item 1 from file: 348)  
DIALOG(R)File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01372102

Apparatus for retrieving similar documents and apparatus for extracting relevant keywords

Gerat zum Wiederauffinden von ahnlichen Dokumenten und Gerat zum Extrahieren von relevanten Schlussselwortern

Appareil de recouvrement de documents similaires et appareil pour extraire des mots-cles pertinents

PATENT ASSIGNEE:

MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD., (216883), 1006, Oaza-Kadoma, Kadoma-shi, Osaka 571-8501, (JP), (Applicant designated States: all)

INVENTOR:

Kanno, Yuji, 1-1-19-504, Oooka, Minami-ku, Yokohama 232-0061, (JP)

LEGAL REPRESENTATIVE:

Dempster, Benjamin John Naftel et al (62251), Withers & Rogers, Goldings House, 2 Hays Lane, London SE1 2HW, (GB)

PATENT (CC, No, Kind, Date): EP 1168202 A2 020102 (Basic)

APPLICATION (CC, No, Date): EP 2001305637 010628;

PRIORITY (CC, No, Date): JP 2000195075 000628

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

ABSTRACT EP 1168202 A2

After three kinds of data, i.e., a keyword frequency-of-appearance (103), a document length (105), and a keyword weight (107) are produced, a document profile vector (111) and a keyword profile vector (109) are calculated. Then, by independently performing the weighted principal component analysis (112,114) considering the document length and the keyword weight, a document feature vector and a keyword feature vectors are obtained. Then, documents and keywords having higher similarity to the feature vectors calculated with reference to the retrieval and extracting conditions are obtained and displayed.

ABSTRACT WORD COUNT: 88

NOTE:

Figure number on first page: 1

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 020102 A2 Published application without search report

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200201	2952
SPEC A	(English)	200201	12559
Total word count - document A			15511
Total word count - document B			0
Total word count - documents A + B			15511

2/5/3 (Item 2 from file: 348)  
DIALOG(R)File 348:EUROPEAN PATENTS  
(c) 2003 European Patent Office. All rts. reserv.

01317593

VECTOR INDEX CREATING METHOD, SIMILAR VECTOR SEARCHING METHOD, AND DEVICES FOR THEM

VEKTORINDEX ERZEUGUNGSVERFAHREN, AHNLICHE VEKTORSUCHVERFAHREN UND GERATE DAFUR

CREATION D'UN INDICE VECTORIEL, RECHERCHE DE VECTEURS SIMILAIRES ET DISPOSITIFS CORRESPONDANTS

PATENT ASSIGNEE:

MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD., (216885), 1006, Oaza Kadoma, Kadoma-shi, Osaka 571-0050, (JP), (Applicant designated States: all)

INVENTOR:

KANNO, Yuji , 504, 1-19, Oooka 1-chome, Minami-ku, Yokohama-shi,  
Kanagawa 232-0061, (JP

LEGAL REPRESENTATIVE:

Casalonga, Axel (14511), BUREAU D.A. CASALONGA - JOSSE Morassistrasse 8,  
80469 Munchen, (DE)

PATENT (CC, No, Kind, Date): EP 1204032 A1 020508 (Basic)

WO 200146858 010628

APPLICATION (CC, No, Date): EP 2000987677 001221; WO 2000JP9079 001221

PRIORITY (CC, No, Date): JP 99363058 991221

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI;  
LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

CITED REFERENCES (WO A):

LEE J.Y. ET AL.: 'A fast retrieval method for image features' TENCON 99.  
PROCEEDINGS OF THE IEEE REGION 10 CONFERENCE vol. 1, 15 September 1999,  
pages 577 - 580

NENE S.A. & NAYAR S.K.: 'A simple algorithm for nearest neighbor search  
in high dimensions' IEEE TRANSACTIONS ON PATTERN ANALYSIS AND  
INTELLIGENCE vol. 19, no. 9, September 1997, pages 989 - 1003

WASHIZAWA: 'k-Nearest neighbor tansaku no kousoku keisan algorithm' GAZOU  
DENSHI GAKKAISHI vol. 27, no. 5, 25 October 1998, JAPAN, pages 513 -  
520

KATAYAMA N. & SATOH S.: 'The SR-tree: An index structure for  
high-dimensional nearest neighbor queries' PROCEEDINGS OF THE ACM  
SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA 11 May 1997,  
TUCSON, AZ, USA, pages 369 - 380

BRIN S.: 'Near neighbor search in large metric spaces' PROCEEDINGS OF THE  
21ST INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES 11 September  
1995, ZURICH, SWITZERLAND, pages 574 - 584

BERCHTOLD S. ET AL.: 'The pyramid-technique: Towards breaking the curse  
of dimensionality' PROCEEDINGS OF ACM SIGMOD INTERNATIONAL CONFERENCE  
ON MANAGEMENT OF DATA 01 June 1998, SEATTLE, WA, USA, pages 142 - 153;

ABSTRACT EP 1204032 A1

In the present invention, a similar vector is searched from a several  
hundreds dimensional vector database at a high speed, by a single **vector**  
**index** , and in accordance with either measure of an inner product or a  
distance by designating a similarity search range and maximum obtained  
pieces number, **vector index** preparation is performed by decomposing  
each vector into a plurality of partial vectors and characterizing the  
vector by a norm division, belonging region and declination division to  
prepare an index, and similarity search is performed by obtaining a  
partial query vector and partial search range from a query vector and  
search range, performing similarity search in each partial space to  
accumulate a difference from the search range and to obtain an upper  
limit value, and obtaining a correct measure from a higher upper limit  
value to obtain a final similarity search result.

ABSTRACT WORD COUNT: 144

NOTE:

Figure number on first page: 1

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 010822 A1 International application. (Art. 158(1))

Application: 010822 A1 International application entering European  
phase

Application: 020508 A1 Published application with search report

Examination: 020508 A1 Date of request for examination: 20011227

LANGUAGE (Publication,Procedural,Application): English; English; Japanese

FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200219	5148
SPEC A	(English)	200219	14938
Total word count - document A			20086
Total word count - document B			0
Total word count - documents A + B			20086

2/5/4 (Item 1 from file: 349)  
DIALOG(R) File 349:PCT FULLTEXT  
(c) 2003 WIPO/Univentio. All rts. reserv.

00813237 \*\*Image available\*\*  
VECTOR INDEX CREATING METHOD, SIMILAR VECTOR SEARCHING METHOD, AND  
DEVICES FOR THEM  
CREATION D'UN INDICE VECTORIEL, RECHERCHE DE VECTEURS SIMILAIRES ET  
DISPOSITIFS CORRESPONDANTS

Patent Applicant/Assignee:

MATSUSHITA ELECTRIC INDUSTRIAL CO LTD, 1006, Oaza Kadoma, Kadoma-shi,  
Osaka 571-0050, JP, JP (Residence), JP (Nationality), (For all  
designated states except: US)

Patent Applicant/Inventor:

KANNO Yuji, 504, 1-19, Oooka 1-chome, Minami-ku, Yokohama-shi, Kanagawa  
232-0061, JP, JP (Residence), JP (Nationality), (Designated only for:  
US

Legal Representative:

NIHEI Masayuki (agent), 6th Tomizawa bldg. 6F, 12-5, Yotsuya 2-chome,  
Shibuya-ku, Tokyo 150-0001, JP,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200146858 A1 20010628 (WO 0146858)

Application: WO 2000JP9079 20001221 (PCT/WO JP0009079)

Priority Application: JP 99363058 19991221

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ

DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ

LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG

SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

Publication Language: Japanese

Filing Language: Japanese

#### English Abstract

Similar vectors are fast retrieved from a vector database of several  
hundreds of dimensions with reference to a single **vector index**  
according to the criterion of the inner product or distance, after  
specifying the similarity search range and the maximum number of similar  
vectors to be retrieved. For creating the **vector index**, each **vector**  
is decomposed into sub-vectors and featured by a norm section, an  
assigned area, and an argument section. For similarity search, a  
sub-query vector and a sub search range are determined from the query  
vector and the search range, similarity search in sub-space is carried  
out, and differences from the search range are cumulated to determine the  
upper limits. An accurate criterion having a higher upper limit is  
preferentially determined, thereby producing a final similarity search  
result.



File 8: Ei Compendex(R) 1970-2003/Sep W2  
(c) 2003 Elsevier Eng. Info. Inc.  
File 35: Dissertation Abs Online 1861-2003/Aug  
(c) 2003 ProQuest Info&Learning  
File 202: Info. Sci. & Tech. Abs. 1966-2003/Sep 16  
(c) 2003 EBSCO Publishing  
File 65: Inside Conferences 1993-2003/Sep W2  
(c) 2003 BLDSC all rts. reserv.  
File 2: INSPEC 1969-2003/Sep W2  
(c) 2003 Institution of Electrical Engineers  
File 94: JICST-EPlus 1985-2003/Sep W3  
(c) 2003 Japan Science and Tech Corp(JST)  
File 6: NTIS 1964-2003/Sep W2  
(c) 2003 NTIS, Intl Cpyrght All Rights Res  
File 144: Pascal 1973-2003/Sep W2  
(c) 2003 INIST/CNRS  
File 434: SciSearch(R) Cited Ref Sci 1974-1989/Dec  
(c) 1998 Inst for Sci Info  
File 34: SciSearch(R) Cited Ref Sci 1990-2003/Sep W2  
(c) 2003 Inst for Sci Info  
File 99: Wilson Appl. Sci & Tech Abs 1983-2003/Aug  
(c) 2003 The HW Wilson Co.  
File 583: Gale Group Globalbase(TM) 1986-2002/Dec 13  
(c) 2002 The Gale Group  
File 266: FEDRIP 2003/Jul  
Comp & dist by NTIS, Intl Copyright All Rights Res  
File 95: TEME-Technology & Management 1989-2003/Aug W5  
(c) 2003 FIZ TECHNIK  
File 438: Library Lit. & Info. Science 1984-2003/Aug  
(c) 2003 The HW Wilson Co  
File 62: SPIN(R) 1975-2003/Aug W2  
(c) 2003 American Institute of Physics  
File 239: Mathsci 1940-2003/Oct  
(c) 2003 American Mathematical Society  
File 248: PIRA 1975-2003/Sep W2  
(c) 2003 Pira International

Set	Items	Description
S1	791587	VECTOR? ?
S2	3730	S1(5N)(INDEX??? OR INDICE? ?)
S3	4391	SIMILARIT???(5N)(SEARCH??? OR QUERIE? ? OR QUERY???)
S4	3890	S1(5N)DECOMPOS?
S5	2685505	IMAGE? ? OR PHOTO? ? OR PHOTOGRAPH? ? OR PICTURE? ?
S6	701591	DATABASE? ? OR DATA()BASE? ? OR REPOSITOR???
S7	77	S2 AND S3
S8	54	S7 AND S4:S6
S9	33	RD (unique items)
S10	26	S9 NOT PY=2002:2003
S11	1519	AU=(KANNO, Y? OR KANNO Y?)
S12	0	S11 AND S2:S4

10/5/1 (Item 1 from file: 8)  
DIALOG(R)File 8: Ei Compendex(R)  
(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

06385268 E.I. No: EIP03207472200

**Title: A vector based approach to color image retrieval**  
Author: Androutsos, D.; Plataniotis, K.N.; Venetsanopoulos, A.N.  
Corporate Source: Digit. Sign./Image Processing Lab Department of  
Electrical Engineering University of Toronto, Toronto, Ont., M5S 3G4,  
Canada

Conference Title: Multimedia Storage and Archiving Systems III  
Conference Location: Boston, MA, United States Conference Date:  
19981102-19981104

Sponsor: SPIE  
E.I. Conference No.: 60962  
Source: Proceedings of SPIE - The International Society for Optical  
Engineering v 3527 1998. p 497-504

Publication Year: 1998  
CODEN: PSISDG ISSN: 0277-786X

Language: English  
Document Type: CA; (Conference Article) Treatment: T; (Theoretical)  
Journal Announcement: 0305W4

Abstract: In this paper we present a novel technique for **image**  
retrieval based on color. Our system is based on color segmentation where  
only a small number of representative color vectors are extracted from  
each **image** and used to build **image indices**. These **vectors** are then  
used with vector distance measures to determine **similarity** between a  
**query** color and a **database image**. We test numerous popular vector  
distance measures in our system, along with popular histogram techniques,  
and find that angular directional measures using our technique provide  
more accurate and perceptually relevant retrievals. 10 Refs.

Descriptors: **Image** retrieval; Color **image** processing; **Image**  
segmentation; **Database** systems; Distance measurement; Vectors

Identifiers: **Image** indices

Classification Codes:

903.3 (Information Retrieval & Use); 723.5 (Computer Applications);  
741.1 (Light & Optics); 723.2 (Data Processing); 723.3 (Database  
Systems); 943.2 (Mechanical Variables Measurements); 921.1 (Algebra)  
903 (Information Science); 723 (Computer Software, Data Handling &  
Applications); 741 (Light, Optics & Optical Devices); 943 (Mechanical &  
Miscellaneous Measuring Instruments); 921 (Applied Mathematics)  
90 (ENGINEERING, GENERAL); 72 (COMPUTERS & DATA PROCESSING); 74 (LIGHT  
& OPTICAL TECHNOLOGY); 94 (INSTRUMENTS & MEASUREMENT); 92 (ENGINEERING  
MATHEMATICS)

10/5/2 (Item 2 from file: 8)  
DIALOG(R)File 8: Ei Compendex(R)  
(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

06385237 E.I. No: EIP03207472169

**Title: CSVD: Approximate similarity searches in high dimensional  
spaces using clustering and singular value decomposition**

Author: Thomasian, Alexander; Castelli, Vittorio; Li, Chung-Sheng  
Corporate Source: IBM Thomas J. Watson Research Center, Yorktown Heights,  
NY 10598, United States

Conference Title: Multimedia Storage and Archiving Systems III  
Conference Location: Boston, MA, United States Conference Date:  
19981102-19981104

Sponsor: SPIE  
E.I. Conference No.: 60962  
Source: Proceedings of SPIE - The International Society for Optical  
Engineering v 3527 1998. p 144-154

Publication Year: 1998  
CODEN: PSISDG ISSN: 0277-786X

Language: English  
Document Type: CA; (Conference Article) Treatment: T; (Theoretical); X;  
(Experimental)

Journal Announcement: 0305W4

**Abstract:** Many data-intensive applications, such as content-based retrieval of **images** or video from multimedia **databases** and similarity retrieval of patterns in data mining, require the ability of efficiently performing **similarity queries**. Unfortunately, the performance of nearest neighbor (NN) algorithms, the basis for **similarity search**, quickly deteriorates with the number of dimensions. In this paper we propose a method called Clustering with Singular Value Decomposition (CSVD), combining clustering and singular value decomposition (SVD) to reduce the number of index dimensions. With CSVD, points are grouped into clusters that are more amenable to dimensionality reduction than the original dataset. Experiments with texture vectors extracted from satellite **images** show that CSVD achieves significantly higher dimensionality reduction than SVD alone for the same fraction of total variance preserved. Conversely, for the same compression ratio CSVD results in an increase in preserved total variance with respect to SVD (e.g., a 70% increase for a 20:1 compression ratio). Then, approximate NN queries are more efficiently processed, as quantified through experimental results. 28 Refs.

**Descriptors:** Pattern matching; Content based retrieval; Multimedia systems; Feature extraction; **Indexing** (of information); Query languages; **Vector** quantization; Discrete Fourier transforms; Cosine transforms; Principal component analysis

**Identifiers:** Singular value decomposition (SVD) methods; Clustering methods

**Classification Codes:**

723.5 (Computer Applications); 723.2 (Data Processing); 903.1 (Information Sources & Analysis); 723.3 (Database Systems); 921.1 (Algebra); 921.3 (Mathematical Transformations); 922.2 (Mathematical Statistics)

723 (Computer Software, Data Handling & Applications); 903 (Information Science); 921 (Applied Mathematics); 922 (Statistical Methods)

72 (COMPUTERS & DATA PROCESSING); 90 (ENGINEERING, GENERAL); 92 (ENGINEERING MATHEMATICS)

10/5/3 (Item 3 from file: 8)

DIALOG(R) File 8: Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

05918433 E.I. No: EIP01436700141

**Title:** On the effective clustering of multidimensional data sequences

**Author:** Lee, S.-L.; Chung, C.-W.

**Corporate Source:** Department of Computer Science Korea Adv. Inst. of Science/Technol., Yusong-Gu, Taejon 305-701, South Korea

**Source:** Information Processing Letters v 80 n 2 Oct 31 2001. p 87-95

**Publication Year:** 2001

**CODEN:** IFPLAT **ISSN:** 0020-0190

**Language:** English

**Document Type:** JA; (Journal Article) **Treatment:** T; (Theoretical)

**Journal Announcement:** 0110W4

**Abstract:** In this paper, we investigate the problem of clustering multidimensional data sequences such as video streams. Each sequence is represented by a small number of hyper-rectangular clusters for subsequent indexing and **similarity search** processing. We present a linear clustering algorithm that guarantees the predefined level of clustering quality, and show its effectiveness via experiments on various video data sets. copy 2001 Elsevier Science B.V. All rights reserved. 9 Refs.

**Descriptors:** **Database** systems; Video signal processing; **Indexing** (of information); **Vectors**; Data reduction; Algorithms

**Identifiers:** Multidimensional data sequences (MDS)

**Classification Codes:**

723.3 (Database Systems); 716.4 (Television Systems & Equipment); 903.1 (Information Sources & Analysis); 921.1 (Algebra); 723.2 (Data Processing)

723 (Computer Software, Data Handling & Applications); 716 (Electronic Equipment, Radar, Radio & Television); 903 (Information Science); 921 (Applied Mathematics)

72 (COMPUTERS & DATA PROCESSING); 71 (ELECTRONICS & COMMUNICATION  
ENGINEERING); 90 (ENGINEERING, GENERAL); 92 (ENGINEERING MATHEMATICS)

10/5/4 (Item 4 from file: 8)  
DIALOG(R)File 8: Ei Compendex(R)  
(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

05844733 E.I. No: EIP01266561392

**Title: An index-based approach for similarity search supporting time warping in large sequence databases**

Author: Kim, S.-W.; Park, S.; Chu, W.W.

Corporate Source: Department of Computer Engineering Kangwon National University, Kangwon, South Korea

Conference Title: 17th International Conference on Data Engineering

Conference Location: Heidelberg, Germany Conference Date: 20010402-20010406

Sponsor: IEEE; EML; IBM; SAS

E.I. Conference No.: 58203

Source: Proceedings - International Conference on Data Engineering 2001.  
p 607-614

Publication Year: 2001

CODEN: PIDEEG

Language: English

Document Type: CA; (Conference Article) Treatment: T; (Theoretical); X; (Experimental)

Journal Announcement: 0107W1

**Abstract:** This paper proposes a new novel method for **similarity search** that supports time warping in large sequence **databases**. Time warping enables finding sequences with similar patterns even when they are of different lengths. Previous methods for processing **similarity search** that supports time warping fail to employ multi-dimensional indexes without false dismissal since the time warping distance does not satisfy the triangular inequality. Our primary goal is to innovate on search performance without permitting any false dismissal. To attain this goal, we devise a new distance function  $D//t//w//l//b$  that consistently underestimates the time warping distance and also satisfies the triangular inequality.  $D//t//w//l//b$  uses a 4-tuple feature vector that is extracted from each sequence and is invariant to time warping. For efficient processing of **similarity search**, we employ a multi-dimensional index that uses the 4-tuple feature **vector** as **indexing** attributes and  $D//t//w//l//b$  as a distance function. The extensive experimental results reveal that our method achieves significant speedup up to 43 times with real-world S&P 500 stock data and up to 720 times with very large synthetic data. 22 Refs.

**Descriptors:** **Database** systems; Data mining; Vectors; Algorithms; Computational geometry

**Identifiers:** Data sequences

**Classification Codes:**

723.3 (Database Systems); 723.2 (Data Processing); 921.1 (Algebra); 723.5 (Computer Applications); 921.4 (Combinatorial Mathematics, Includes Graph Theory, Set Theory)

723 (Computer Software, Data Handling & Applications); 921 (Applied Mathematics)

72 (COMPUTERS & DATA PROCESSING); 92 (ENGINEERING MATHEMATICS)

10/5/5 (Item 5 from file: 8)  
DIALOG(R)File 8: Ei Compendex(R)  
(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

05299732 E.I. No: EIP99064700827

**Title: Query vector projection access method**

Author: Smith, John R.

Corporate Source: IBM T.J. Watson Research Cent, Hawthorne, NY, USA

Conference Title: Proceedings of the 1999 7th Conference of the Storage and Retrieval for Image and Video Databases VII

Conference Location: San Jose, Ca, USA Conference Date:

19990126-19990129

Sponsor: Is and T; SPIE

E.I. Conference No.: 55094

Source: Proceedings of SPIE - The International Society for Optical Engineering v 3656 1999. p 511-522

Publication Year: 1999

CODEN: PSISDG ISSN: 0277-786X

Language: English

Document Type: JA; (Journal Article) Treatment: T; (Theoretical)

Journal Announcement: 9908W1

Abstract: We present a new multi-dimensional access method for **querying** by **similarity** in **databases** of high-dimensional vectors. The query vector projection access method (QVPAM) addresses the shortcomings of other dimensionality reduction techniques by deriving the best transformation of the vectors at query-time. QVPAM creates a projection library that contains building blocks for constructing the transformations. QVPAM rapidly searches the projection library at query-time in order to select the set of projection elements that minimizes the work for processing the query. Since the selected set does not need to be complete, QVPAM effectively trades-off query precision and query response time. We describe QVPAM and demonstrate its performance in the content-based querying of a **database** of high-dimensional color histograms. (Author abstract) 15 Refs.

Descriptors: **Database** systems; Data acquisition; Multimedia systems; Vectors; Information retrieval; Mathematical transformations; Response time (computer systems); Indexing (of information); Color **image** processing

Identifiers: Query **vector** projection access method; Multidimensional **indexing**; **Similarity** **search**; Content based **query**

Classification Codes:

723.3 (Database Systems); 723.2 (Data Processing); 723.5 (Computer Applications); 921.1 (Algebra); 903.3 (Information Retrieval & Use); 921.3 (Mathematical Transformations)

723 (Computer Software); 921 (Applied Mathematics); 903 (Information Science)

72 (COMPUTERS & DATA PROCESSING); 92 (ENGINEERING MATHEMATICS); 90 (GENERAL ENGINEERING)

10/5/6 (Item 6 from file: 8)

DIALOG(R)File 8:Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

05277414 E.I. No: EIP99054666396

Title: **Downdating the latent semantic indexing model for conceptual information retrieval**

Author: Witter, Dian I.; Berry, Michael W.

Corporate Source: Univ of Tennessee, Knoxville, TN, USA

Source: Computer Journal v 41 n 8 1998. p 589-601

Publication Year: 1998

Language: English

Document Type: JA; (Journal Article) Treatment: T; (Theoretical)

Journal Announcement: 9906W5

Abstract: Due to the growth of large data collections, information retrieval or **database** searching is of vital importance. Lexical matching techniques may retrieve irrelevant or inaccurate results because of synonyms and polysemous words, so effective concept-based techniques are needed. One such technique is latent semantic **indexing** (LSI) which uses a **vector**-space approach by identifying documents whose content is related to the user's **query** in order of **similarity**. LSI uses the singular value decomposition (SVD) of term-by-document matrix to encode the terms and documents in a vector-space model. Existing methods for removing terms or documents from the term-document space are either time consuming or do not sufficiently change the term-document relationships. This paper presents a new method for downdating, downdating the reduced model (or DRM) method, and discusses its implementation into the LSI plus plus software environment. The DRM method can be used to assess the effect that a term or document has on the clustering of relevant information in a collection and for the incorporation of user feedback in the existing LSI model. Implementing the DRM method within LSI plus plus not only provides

downdating functionality, but is less time consuming than recomputing the SVD when removing a term, document or both. The DRM method is a viable algorithm for dynamic information modeling and retrieval. (Author abstract) 20 Refs.

Descriptors: Information retrieval; **Database** systems; Computational linguistics; **Indexing** (of information); **Vectors** ; Matrix algebra; Encoding (symbols); Computer software; Computer simulation; Statistical methods

Identifiers: Latent semantic indexing (LSI); Singular value decomposition (SVD); Downdating the reduced model (DRM) methods

Classification Codes:

903.3 (Information Retrieval & Use); 723.5 (Computer Applications); 723.3 (Database Systems); 721.1 (Computer Theory, Includes Formal Logic, Automata Theory, Switching Theory, Programming Theory); 903.1 (Information Sources & Analysis); 921.1 (Algebra)

903 (Information Science); 723 (Computer Software); 721 (Computer Circuits & Logic Elements); 921 (Applied Mathematics)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING); 92 (ENGINEERING MATHEMATICS)

10/5/7 (Item 7 from file: 8)

DIALOG(R)File 8: Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

04372764 E.I. No: EIP96043118379

**Title: Similarity indexing with the SS-tree**

Author: White, David A.; Jain, Ramesh

Corporate Source: Univ of California, San Diego, La Jolla, CA, USA

Conference Title: Proceedings of the 1996 IEEE 12th International Conference on Data Engineering

Conference Location: New Orleans, LA, USA Conference Date: 19960226-19960301

Sponsor: IEEE

E.I. Conference No.: 44487

Source: Proceedings - International Conference on Data Engineering 1996. IEEE, Los Alamitos, CA, USA, 96CB35888. p 516-523

Publication Year: 1996

CODEN: PIDEEG

Language: English

Document Type: CA; (Conference Article) Treatment: G; (General Review); T; (Theoretical)

Journal Announcement: 9605W4

Abstract: Efficient **indexing** of high dimensional feature **vectors** is important to allow visual information systems and a number other applications to scale up to large **databases** . In this paper, we define this problem as 'similarity indexing' and describe the fundamental types of 'similarity queries' that we believe should be supported. We also propose a new dynamic structure for **similarity** indexing called the **similarity search** tree or SS-tree. In nearly every test we performed on high dimensional data, we found that this structure performed better than the R\*-tree. Our tests also show that the SS-tree is much better suited for approximate queries than the R\*-tree. (Author abstract) 28 Refs.

Descriptors: \*Indexing (of information); Query languages; Trees (mathematics); Information retrieval systems; Data structures

Identifiers: **Similarity** indexing; **Similarity search** tree

Classification Codes:

903.1 (Information Sources & Analysis); 723.3 (Database Systems); 921.4 (Combinatorial Mathematics, Includes Graph Theory, Set Theory); 903.3 (Information Retrieval & Use); 723.2 (Data Processing)

903 (Information Science); 723 (Computer Software); 921 (Applied Mathematics)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING); 92 (ENGINEERING MATHEMATICS)

10/5/8 (Item 8 from file: 8)

DIALOG(R)File 8: Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

03939608 E.I. No: EIP94091397560

**Title: Query-based mechanism for geometrical objects retrieval in multimedia information system**

Author: Davcev, Danco; Cakmakov, Dusan; Arnautovska, Vesna

Corporate Source: 'Kiril & Metodij' Univ, Skopje, Macedonia

Conference Title: Proceedings of the 27th Hawaii International Conference on System Sciences (HICSS-27). Part 3 (of 5)

Conference Location: Wailea, HI, USA Conference Date: 19940104-19940107

Sponsor: University of Hawaii; University of Hawaii College of Business Administration; IEEE Computer Society; Association for Computing Machinery  
E.I. Conference No.: 20790

Source: Proceedings of the Hawaii International Conference on System Sciences v 3 1994. Publ by IEEE, Computer Society Press, Los Alamitos, CA, USA, 94TH0607-2. p 581-589

Publication Year: 1994

CODEN: PHISD7 ISSN: 1060-3425 ISBN: 0-8186-5070-2

Language: English

Document Type: CA; (Conference Article) Treatment: G; (General Review); T; (Theoretical)

Journal Announcement: 9410W4

**Abstract:** The basic elements of A Multimedia Cognitive-based Information Retrieval System called AMCIRS which integrates **image** and text information have been described elsewhere. The AMCIRS query based mechanism is based on multimedia objects content search using the vector model. The content **search** process is deduced to the **similarity** estimation between **query** and **index vectors**. The main objective of this paper is to introduce the similarity estimation model for geometrical objects as a part of a query mechanism of AMCIRS. Our model for polygon similarity estimation introduces a numerical measure of similarity between two polygons and gives acceptable results for all polygon forms and any number of vertices. The algorithm based on this model as well as the simulation results are also given. (Author abstract) 27 Refs.

**Descriptors:** \*Information retrieval systems; Query languages; Information retrieval; Information services; Mathematical models; Geometry; Vectors; Algorithms; Information management; Fourier transforms

**Identifiers:** Query based mechanisms; Geometrical objects; A multimedia cognitive based information retrieval system (AMCIRS); Vector model

**Classification Codes:**

903.3 (Information Retrieval & Use); 722.4 (Digital Computers & Systems); 723.3 (Database Systems); 903.4 (Information Services); 921.6 (Numerical Methods); 921.3 (Mathematical Transformations)

903 (Information Science); 722 (Computer Hardware); 723 (Computer Software); 921 (Applied Mathematics)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING); 92 (ENGINEERING MATHEMATICS)

10/5/9 (Item 9 from file: 8)

DIALOG(R)File 8:Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

03854533 E.I. No: EIP94051281502

**Title: Application of a multimedia cognitive-based information retrieval system (AMCIRS) in mineralogy**

Author: Davcev, Danco; Cakmakov, Dusan

Corporate Source: 'Kiril & Metodij' Univ, Skopje, Macedonia

Conference Title: Proceedings of the 21st Annual Computer Science Conference

Conference Location: Indianapolis, IN, USA Conference Date: 19930216-19930218

E.I. Conference No.: 20238

Source: Proceedings - ACM Computer Science Conference 1993. Publ by ACM, New York, NY, USA. p 284-290

Publication Year: 1993

CODEN: PCSCE9 ISBN: 0-89791-558-5

Language: English

Document Type: CA; (Conference Article) Treatment: A; (Applications); G  
; (General Review); T; (Theoretical)

Journal Announcement: 9406W3

Abstract: A Multimedia Cognitive-based Information Retrieval System called AMCIRS which integrates **image** and text information has been described in left bracket 11 right bracket , left bracket 12 right bracket . The AMCIRS query based mechanism is based on multimedia objects content search using the vector model. The content **search** process is deduced to the **similarity** estimation between **query** and **index vectors** . The main objective of this paper is to present an application of AMCIRS in Mineralogy. The experimental evaluation of AMCIRS retrieval effectiveness is also given. The retrieval effectiveness is expressed by recall and precision parameters which are the standard measures for the effectivity of the information retrieval systems. We confirmed our assumption that multiple media retrieval has advantages with respect to single media retrieval. (Author abstract) 26 Refs.

Descriptors: Information retrieval systems; Cognitive systems; **Image** processing; Data processing; Query languages; Associative storage; Mathematical models; Vectors; Estimation; Mineralogy

Identifiers: Multimedia cognitive based information retrieval system (AMCIRS); Multimedia object content search; **Vector** model; **Index** vectors ; **Query vectors** ; **Similarity** estimation; Retrieval effectiveness

Classification Codes:

903.3 (Information Retrieval & Use); 723.4 (Artificial Intelligence); 723.2 (Data Processing); 741.2 (Vision); 723.3 (Database Systems); 722.1 (Data Storage, Equipment & Techniques)

903 (Information Science); 723 (Computer Software); 741 (Optics & Optical Devices); 722 (Computer Hardware)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING); 74 (OPTICAL TECHNOLOGY)

10/5/10 (Item 10 from file: 8)

DIALOG(R)File 8: Ei Compendex(R)

(c) 2003 Elsevier Eng. Info. Inc. All rts. reserv.

03790882 E.I. No: EIP94011196077

**Title: Experiments in retrieval of mineral information**

Author: Cakmakov, Dusan; Davcev, Danco

Corporate Source: ?Kiril & Metodij' Univ, Skopje, Macedonia

Conference Title: Proceedings of the 1st ACM International Conference on Multimedia

Conference Location: Anaheim, CA, USA Conference Date: 19930801-19930806

Sponsor: ACM, SIGBIO; ACM, SIGCHI; ACM, SIGCOMM; ACM, SIGGRAPH; ACM, SIGIR; et al

E.I. Conference No.: 19832

Source: Proc 1 ACM Int Conf Multimedia 1993. Publ by ACM, New York, NY, USA. p 57-64

Publication Year: 1993

ISBN: 0-89791-596-8

Language: English

Document Type: CA; (Conference Article) Treatment: A; (Applications); T ; (Theoretical); X; (Experimental)

Journal Announcement: 9403W2

Abstract: In this paper, we present the experiments in retrieval of multimedia mineral information using AMCIRS (A Multimedia Cognitive-based Information Retrieval System). The AMCIRS query based mechanism is based on a multimedia objects content search using the vector model. Each vector is composed of text and **image** objects. The **image** objects in the vectors are **image** object contours, represented by polygonal approximations. The content **search** process is deduced to the **similarity** estimation between the MM **query** and MM **index vectors** . The **similarity** function for **image** objects is based on the polygon similarity estimation. The experimental evaluation of AMCIRS retrieval effectiveness is expressed by the recall and precision parameters. Possible advantages of multiple media retrieval with respect to the single medium retrieval are also investigated and explicitly represented by the recall-precision diagrams. (Author



abstract) 27 Refs.

Descriptors: \*Information retrieval systems; Cognitive systems; Minerals; Online searching; Query languages; Mathematical models; Information analysis; Knowledge based systems; Parameter estimation; Approximation theory

Identifiers: Multimedia cognitive based information retrieval system; Polygon **similarity** estimation; Content **search**; Information recall; Precision

Classification Codes:

903.3 (Information Retrieval & Use); 723.5 (Computer Applications); 482.2 (Minerals); 723.3 (Database Systems); 903.2 (Information Dissemination)

903 (Information Science); 723 (Computer Software); 482 (Mineralogy & Petrology); 922 (Statistical Methods)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING); 48 (ENGINEERING GEOLOGY); 92 (ENGINEERING MATHEMATICS)

10/5/11 (Item 1 from file: 35)

DIALOG(R)File 35:Dissertation Abs Online

(c) 2003 ProQuest Info&Learning. All rts. reserv.

01771492 ORDER NO: AADAA-IC801556

Image **texture analysis with fast similarity search for content-based retrieval and navigation**

Author: Kuan, Joseph

Degree: Ph.D.

Year: 1998

Corporate Source/Institution: University of Southampton (United Kingdom) (5036)

Source: VOLUME 61/02-C OF DISSERTATION ABSTRACTS INTERNATIONAL.

PAGE 593

Descriptors: ENGINEERING, ELECTRONICS AND ELECTRICAL

Descriptor Codes: 0544

One of the main challenges of multimedia and hypermedia research is the effective use of the media content for retrieval and navigation in multimedia environments. This thesis is concerned with the use of texture as one of the keys for content based retrieval (CBR) and content based navigation (CBN). Other authors have proposed texture analysis procedures and an initial aim was to identify a versatile texture representation which is effective over a very wide range of textures and which could be used efficiently in the context of CBR and CBN. In order to **index** the multidimensional feature **vectors** representing texture efficiently, this thesis has also focused on issues of multidimensional indexing for fast **similarity search**.

This thesis proposes a novel texture representation method which uses the edge and plain region information from texture patterns. The information is used to evaluate contrast across edges, the mean greylevel of plain regions and the conditional probability matrix of edge directions and plain regions as features. A weighted Euclidean measurement for this method is proposed which gives better matching than the standard Euclidean measure. The new representation is compared with a range of previous texture representation schemes using a wide range of texture patterns and its classification properties and speed performance are shown to be an improvement on the other schemes. Since texture is typically represented by a multidimensional feature **vector**, this thesis investigates multidimensional **indexing** strategies and *nn* retrieval methods and proposes new and more efficient approaches in the context of multimedia information handling.

Two different multidimensional indexing approaches are explored in this thesis; the R\*-tree and the Hilbert R-tree. Data object retrieval and range search performance are compared in various aspects, including the number of dimensions, nature of **databases** and **database** size. *k* nearest neighbours (*k* *nn*) **similarity search** is significant for **image** based CBR and CBN in multimedia systems, and a new algorithm for *k* *nn* search is proposed which is an improvement over previous approaches for **image** based CBR and CBN

applications.

The novel texture representation technique, fast indexing R-tree and the enhanced R-tree **similarity search** technique are integrated into an open hypermedia system which offers content based retrieval and navigation for multimedia data. The thesis concludes with examples of the use of the system for texture based **image** retrieval and also texture based navigation from **images** to other media types.

10/5/12 (Item 2 from file: 35)  
DIALOG(R)File 35:Dissertation Abs Online  
(c) 2003 ProQuest Info&Learning. All rts. reserv.

01614180 ORDER NO: AAD98-12309

**A FEATURE-BASED INDEXING FOR SPATIAL DATA OBJECTS**

Author: YAZDANI, NASSER

Degree: PH.D.

Year: 1997

Corporate Source/Institution: CASE WESTERN RESERVE UNIVERSITY (0042)

Adviser: Z. MERAL OZSOYUGLU

Source: VOLUME 58/10-B OF DISSERTATION ABSTRACTS INTERNATIONAL.

PAGE 5493. 176 PAGES

Descriptors: COMPUTER SCIENCE

Descriptor Codes: 0984

We propose a feature-based indexing method for spatial data objects. The aim is to efficiently retrieve the data objects as well as similar objects with a given query object in a spatial **database** environment. Our method extracts some features from each data object in order to build an index tree. A broad range of problems and issues such as indexing and modeling, similarity matching, transformations, features and spatial access methods must be dealt with in any feature-based indexing method. Our work consists of two parts. First, we propose a framework for feature-based indexing of **image** data and apply our method to the damage zone shapes of materials. A set of generic features which are invariant to rotation, translation and scaling for the sake of similarity matching are proposed. These features form a feature vector for each **image**. The feature vectors are extended with some domain specific features. The feature **vectors** are used to build the **index** structure. Any multi-dimensional point access method can then be used to build the index. However we use a variant of the K-D-B tree. Weighted Euclidean distance is used as similarity measure. Each feature in the feature vector is associated with a weight, based on the application, which is used in the **search** process for **similarity** matching. A formula is proposed to find the similarity of nodes in the index tree with a given query shape. This formula is used to prune the search tree in the query processing.

In the second part, we propose two inter-sequence matching methods for exact and similarity matching of **image** sequences. We assume that the **image** sequence matching problem is transformed into matching sequences of real numbers. The methods do not require sequences to have the same length. The first method tries to find the Longest Matching Subsequences (LMS) of two sequences to be matched and uses a modified version of the Longest Common Subsequence (LCS) method for actual matching. In the second method, a modified version of restricted edit distance is used for matching. We also propose a feature-based indexing mechanism to filter out those sequences which are matching candidates with a given query sequence from a large data set. Like all other feature-based indexing methods, our method maps each sequence into a point in K dimensional space, where K is the number of extracted features for the sequence. It operates in two phases, hypothesizing and verification. Lengths and moments (mean and variance) of sequences are used as features. Experimental results indicate that the features and proposed method for query processing perform well as a filter.

10/5/13 (Item 1 from file: 2)  
DIALOG(R)File 2:INSPEC  
(c) 2003 Institution of Electrical Engineers. All rts. reserv.

7491991 INSPEC Abstract Number: B2003-02-7510-010, C2003-02-7330-126

**Title:** Medical image indexing and compression based on vector quantization: image retrieval efficiency evaluation

**Author(s):** Ordenez, J.R.; Cazuguel, G.; Puentes, J.; Solaiman, B.; Cauvin, J.M.; Roux, C.

**Author Affiliation:** INSERM, Brest, France

**Conference Title:** 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.01CH37272) Part vol.3 p.2465-8 vol.3

**Publisher:** IEEE, Piscataway, NJ, USA

**Publication Date:** 2001 **Country of Publication:** USA 4 vol. 4132 pp.

**ISBN:** 0 7803 7211 5 **Material Identity Number:** XX-2002-02147

**U.S. Copyright Clearance Center Code:** 0-7803-7211-5/01/\$17.00

**Conference Title:** 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society

**Conference Date:** 25-28 Oct. 2001 **Conference Location:** Istanbul, Turkey

**Medium:** Also available on CD-ROM in PDF format

**Language:** English **Document Type:** Conference Paper (PA)

**Treatment:** Practical (P); Theoretical (T); Experimental (X)

**Abstract:** Addresses the problem of efficient **image** retrieval from a compressed **image database**, using information derived from the compression process. **Images** in the **database** are compressed applying two approaches: **vector** quantization (VQ) and quadtree **image decomposition**. Both are based on Konohen's self-organizing feature maps (SOFM) for creating vector quantization codebooks. However, while VQ uses one codebook of one resolution to compress the **images**, Quadtree decomposition uses simultaneously 4 codebooks of four different resolutions. **Image** indexing is implemented by generating a feature vector (FV) for each compressed **image**. Accordingly, **images** are retrieved by means of FVs **similarity** evaluation between the **query image** and the **images** in the **database**, depending on a distance measure. Three distance measures have been analyzed to assess FV index similarity: Euclidean, intersection and correlation distances. Distance measures efficiency retrieval is evaluated for different VQ resolutions and different quadtree **image** descriptors. Experimental results using real data, esophageal ultrasound and eye angiography **images**, are presented. (8 Refs)

**Subfile:** B C

**Descriptors:** biomedical ultrasonics; eye; **image** coding; **image** retrieval; medical **image** processing; quadtrees; self-organising feature maps; vector quantisation

**Identifiers:** medical **image** indexing; medical **image** compression; **image** retrieval efficiency evaluation; compressed **image database**; vector quantization; quadtree **image** decomposition; Konohen self-organizing feature maps; vector quantization codebooks; resolution; feature vector; query **image**; similarity evaluation; distance measure; feature **vector index** similarity; Euclidean distances; intersection distances; correlation distances; real data; esophageal ultrasound **images**; eye angiography **images**

**Class Codes:** B7510 (Biomedical measurement and imaging); B6135C (Image and video coding); B7510H (Sonic and ultrasonic radiation (biomedical imaging/measurement)); B0250 (Combinatorial mathematics); C7330 (Biology and medical computing); C5260B (Computer vision and image processing techniques); C5290 (Neural computing techniques); C6160S (Spatial and pictorial databases); C1160 (Combinatorial mathematics)

Copyright 2003, IEE

10/5/14 (Item 2 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

7300728 INSPEC Abstract Number: C2002-07-6160Z-046

**Title:** Prefix-querying: an approach for effective subsequence matching under time warping in sequence databases

**Author(s):** Sanghyun Park; Sang-Wook Kim; June-Suh Cho; Padmanabhan, S.

**Conference Title:** Proceedings of the 2001 ACM CIKM. Tenth International Conference on Information and Knowledge Management p.255-62

Editor(s): Paques, H.; Liu, L.; Grossman, D.  
Publisher: ACM, New York, NY, USA  
Publication Date: 2001 Country of Publication: USA xvii+597 pp.  
ISBN: 1 58113 436 3 Material Identity Number: XX-2002-00537  
U.S. Copyright Clearance Center Code: 1 58113 436 3/2001/0011...\$5.00  
Conference Title: Proceedings of CIKM'01: International Conference on  
Information and Knowledge Management  
Conference Sponsor: ACM SIGIR & SIGMIS  
Conference Date: 5-10 Nov. 2001 Conference Location: Atlanta, GA, USA  
Language: English Document Type: Conference Paper (PA)  
Treatment: Practical (P)

Abstract: This paper discusses an index-based subsequence matching that supports time warping in large sequence **databases**. Time warping enables finding sequences with similar patterns even when they are of different lengths. In our earlier work, we suggested an efficient method for whole matching under time warping. This method constructs a multidimensional **index** on a set of feature **vectors**, which are invariant to time warping, from data sequences. For filtering at feature space, it also applies a lower-bound function, which consistently underestimates the time warping distance as well as satisfies the triangular inequality. We incorporate the prefix-querying approach based on sliding windows into the earlier approach. For **indexing**, we extract a feature **vector** from every subsequence inside a sliding window and construct a multidimensional **index** using a feature **vector** as **indexing** attributes. For query processing, we perform a series of **index** searches using the feature **vectors** of qualifying query prefixes. Our approach provides effective and scalable subsequence matching even with a large volume of a **database**. We also prove that our approach does not incur false dismissal. To verify the superiority of our method, we perform extensive experiments. The results reveal that our method achieves significant speedup with real-world S&P 500 stock data and with very large synthetic data. (21 Refs)

Subfile: C

Descriptors: data mining; **database** indexing; financial data processing; pattern matching; query processing; sequences; very large **databases**

Identifiers: prefix querying; time warping; index-based subsequence matching; large sequence **databases**; pattern recognition; multidimensional index; data sequences; lower-bound function; sliding windows; indexing; feature vector; query processing; index searches; experiments; stock market data; **similarity search**; data mining

Class Codes: C6160Z (Other DBMS); C6170K (Knowledge engineering techniques); C7120 (Financial computing)

Copyright 2002, IEE

10/5/15 (Item 3 from file: 2)  
DIALOG(R)File 2:INSPEC  
(c) 2003 Institution of Electrical Engineers. All rts. reserv.

7289482 INSPEC Abstract Number: C2002-07-6160S-017

Title: An extendible hash for multi-precision similarity querying of image databases

Author(s): Shu Lin; Ozsu, M.T.; Oria, V.; Ng, R.

Author Affiliation: IBM Canada Ltd., Toronto, Ont., Canada

Conference Title: Proceedings of the 27th International Conference on Very Large Data Bases p.221-30

Editor(s): Apers, P.M.G.; Atzeni, P.; Ceri, S.; Paraboschi, S.; Ramamohanarao, K.; Snodgrass, R.T.

Publisher: Morgan Kaufmann Publishing, Orlando, FL, USA

Publication Date: 2001 Country of Publication: USA xxiv+727 pp.

ISBN: 1 55860 804 4 Material Identity Number: XX-2001-02683

Conference Title: Proceedings of Very Large Databases

Conference Sponsor: Inf. Soc. Technol.; Microsoft Res.; IBM; ETNOTEAM; CNR; ROMA TRE: et al

Conference Date: 11-14 Sept. 2001 Conference Location: Rome, Italy

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: We propose multi-precision similarity matching where the **image** is divided into a number of subblocks, each with its associated color

histogram. We present experimental results showing that the spatial distribution information recorded by multiprecision color histograms helps to make similarity matching more precise. We also show that sub- **image** queries are much better supported with multi-precision color histograms. To minimize the overhead, we employ a filtering scheme based on the 3-dimensional average color vectors. We provide a formal result proving that filtering with multi-precision color histograms is complete. Finally, we develop a novel extendible hashing structure for **indexing** the average color **vectors** . We give experimental results showing that the proposed structure significantly outperforms the SR-tree. (21 Refs)

Subfile: C

Descriptors: content-based retrieval; **database** indexing; file organisation; **image** colour analysis; **image** matching; **image** retrieval; visual **databases**

Identifiers: multi-precision similarity matching; multi-precision **similarity** **querying** ; **image** **databases** ; extendible hash; subblocks; spatial distribution information; multi-precision color histograms; sub-**image** queries; filtering scheme; 3D average color vectors; average color **vector** **indexing**

Class Codes: C6160S (Spatial and pictorial databases); C6120 (File organisation); C5260B (Computer vision and image processing techniques)

Copyright 2002, IEE

10/5/16 (Item 4 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

6956049 INSPEC Abstract Number: C2001-07-4250-007

**Title: Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches**

Author(s): Changzhou Wang; Wang, X.S.

Author Affiliation: Dept. of Math. & Comput. Technol., Boeing Co., Bellevue, WA, USA

Journal: VLDB Journal vol.9, no.4 p.344-61

Publisher: Springer-Verlag,

Publication Date: 2001 Country of Publication: Germany

CODEN: VLDBFR ISSN: 1066-8888

SICI: 1066-8888(2001)9:4L:344:IVHD;1-2

Material Identity Number: O851-2001-002

Language: English Document Type: Journal Paper (JP)

Treatment: Theoretical (T)

**Abstract:** **Similarity** **queries** on complex objects are usually translated into searches among their feature **vectors** . The paper studies **indexing** techniques for very high-dimensional (e.g., in hundreds) vectors that are sparse or quasi-sparse, i.e., vectors each having only a small number of non-zero or significant values. Based on the R-tree, the paper introduces the xS-tree that uses lossy compression of bounding regions to guarantee a reasonable minimum fan-out within the allocated storage space for each node. In addition, the paper studies the performance and scalability of the xS-tree via experiments. (29 Refs)

Subfile: C

Descriptors: data compression; **database** indexing; query processing; tree data structures; trees (mathematics); vectors; very large **databases**

Identifiers: very high-dimensional sparse **vector** **indexing** ; quasi-sparse vectors; **similarity** **searches** ; **similarity** **queries** ; complex objects; feature vectors; indexing techniques; significant values; R-tree; xS-tree; lossy compression; bounding regions; minimum fan-out; allocated storage space

Class Codes: C4250 (Database theory); C1110 (Algebra); C1160 (Combinatorial mathematics); C6120 (File organisation); C6130 (Data handling techniques); C6160Z (Other DBMS)

Copyright 2001, IEE

10/5/18 (Item 6 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

6477726 INSPEC Abstract Number: C2000-03-6160M-001

**Title:** Efficient multimedia image retrieval system

**Author(s):** Celenk, M.

**Author Affiliation:** Sch. of Electr. Eng. & Comput. Sci., Ohio Univ., Athens, OH, USA

**Journal:** Proceedings of the SPIE - The International Society for Optical Engineering Conference  
**Title:** Proc. SPIE - Int. Soc. Opt. Eng. (USA)  
vol.3716 p.92-9

**Publisher:** SPIE-Int. Soc. Opt. Eng.,

**Publication Date:** 1999 **Country of Publication:** USA

**CODEN:** PSISDG **ISSN:** 0277-786X

**SICI:** 0277-786X(1999)3716L:92:EMIR;1-M

**Material Identity Number:** C574-1999-255

**U.S. Copyright Clearance Center Code:** 0277-786X/99/\$10.00

**Conference Title:** Visual Information Processing VIII

**Conference Sponsor:** SPIE

**Conference Date:** 6 April 1999 **Conference Location:** Orlando, FL, USA

**Language:** English **Document Type:** Conference Paper (PA); Journal Paper (JP)

**Treatment:** Theoretical (T); Experimental (X)

**Abstract:** This paper describes a computationally efficient method for fast retrieval of color **images** of multimedia and imaging **databases**. Although the proposed algorithm can operate in an n-dimensional feature space for search, in our experiments we use only one 3D **vector** as key for **indexing** and searching color **pictures** of the selected archives. A new feature extraction and matching technique is developed based on the first-order statistics of color **image** data. Eigenvalue analysis provides an effective way of reducing 3-D color data to a one-dimensional (1-D) array. This feature extraction and reduction step is performed only once when an (R,G,B) color **picture** is submitted for storage or **query**. For a **similarity** measure, the Tanimoto coefficient is selected to be a computationally high performance matching algorithm to evaluate the search results. It is shown that the idea of projection-based retrieval is similar to the well-known histogram intersection operation of Swain and Ballard (1991). The algorithm described has been tested on eleven different **databases**, each of which consists of various color **images** of different scenes stored in a content addressable stack. The efficacy of retrieval was determined using the percentage efficiency measure  $\eta = p/P$ , where p is the number of similar **pictures** retrieved in a short list and P is the total number of similar **pictures** in an archive. The experimental results yield almost 90% average retrieval efficiency for the eleven **databases** searched with the 3-D **index** or key **vector**. (16 Refs)

**Subfile:** C

**Descriptors:** content-based retrieval; feature extraction; **image** colour analysis; **image** matching; multimedia **databases**; visual **databases**

**Identifiers:** efficient multimedia **image** retrieval system; computationally efficient method; multimedia **database**; imaging **database**; experiments; 3D vector; color **pictures** indexing; color **pictures** searching; feature extraction; feature matching; first-order statistics; color **image** data; eigenvalue analysis; similarity measure; Tanimoto coefficient; matching algorithm; projection-based retrieval; histogram intersection operation; content addressable stack; percentage efficiency measure; experimental results; key vector; 3D index

**Class Codes:** C6160M (Multimedia databases); C6160S (Spatial and pictorial databases); C7250R (Information retrieval techniques); C5260B (Computer vision and image processing techniques)

Copyright 2000, IEE

10/5/19 (Item 7 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

6346549 INSPEC Abstract Number: C1999-10-6160Z-010

**Title:** An indexing scheme for fast similarity search in large time series databases

**Author(s):** Keogh, E.J.; Pazzani, M.J.

Author Affiliation: Dept. of Inf. & Comput. Sci., California Univ., Irvine, CA, USA

Conference Title: Proceedings. Eleventh International Conference on Scientific and Statistical Database Management p.56-67

Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA

Publication Date: 1999 Country of Publication: USA xiii+287 pp.

ISBN: 0 7695 0046 3 Material Identity Number: XX-1999-02058

U.S. Copyright Clearance Center Code: 0 7695 0046 3/99/\$10.00

Conference Title: Proceedings of Eleventh International Conference on Scientific and Statistical Database Management'99

Conference Sponsor: Case Western Univ.; ACM SIGMOD; VLDB Endowment

Conference Date: 28-30 July 1999 Conference Location: Cleveland, OH, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P); Theoretical (T)

Abstract: Addresses the problem of **similarity searching** in large time-series **databases**. We introduce a novel indexing algorithm that allows faster retrieval. The index is formed by creating bins that contain time series subsequences of approximately the same shape. For each bin, we can quickly calculate a lower bound on the distance between a given query and the most similar element of the bin. This bound allows us to search the bins in best-first order, and to prune some bins from the search space without having to examine the contents. Additional speedup is obtained by optimizing the data within the bins such that we can avoid having to compare the query to every item in the bin. We call our approach STB (Shape To Bit- **vector** ) **indexing**, and experimentally validate it on space telemetry, medical and synthetic data, demonstrating approximately an order-of-magnitude speedup. (25 Refs)

Subfile: C

Descriptors: **database** indexing; **database** theory; medical information systems; query processing; software performance evaluation; space telemetry ; statistical **databases** ; time series; very large **databases**

Identifiers: **similarity searching** ; large time series **databases** ; indexing algorithm; fast retrieval; bins; time series subsequences; lower bound; query distance; best-first search; search space pruning; speedup; data optimization; STB indexing; shape-to-bit **vector indexing** ; space telemetry data; medical data; synthetic data

Class Codes: C6160Z (Other DBMS); C1140Z (Other topics in statistics); C4250 (Database theory)

Copyright 1999, IEE

10/5/20 (Item 8 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

6209132 INSPEC Abstract Number: B1999-05-6135-269, C1999-05-7250R-014

Title: **Distance measures for color image retrieval**

Author(s): Androutsos, D.; Plataniotiss, K.N.; Venetsanopoulos, A.N.

Author Affiliation: Dept. of Electr. & Comput. Eng., Toronto Univ., Ont., Canada

Conference Title: Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269) Part vol.2 p.770-4 vol.2

Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA

Publication Date: 1998 Country of Publication: USA 3 vol. (lxxi+962+984+1013) pp.

ISBN: 0 8186 8821 1 Material Identity Number: XX-1998-02811

U.S. Copyright Clearance Center Code: 0 8186 8821 1/98/\$10.00

Conference Title: Proceedings of IPCIP'98 International Conference on Image Processing

Conference Sponsor: IEEE Signal Process. Soc

Conference Date: 4-7 Oct. 1998 Conference Location: Chicago, IL, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Theoretical (T); Experimental (X)

Abstract: We address the issue of **image database** retrieval based on color using various vector distance metrics. Our system is based on color segmentation where only a few representative color vectors are extracted from each **image** and used as **image indices**. These **vectors** are then

used with vector distance measures to determine **similarity** between a **query** color and a **database** **image** . We test numerous popular vector distance measures in our system and find that directional measures provide the most accurate and perceptually relevant retrievals. (7 Refs)

Subfile: B C

Descriptors: content-based retrieval; feature extraction; **image** colour analysis; **image** segmentation; visual **databases** ; visual perception

Identifiers: color **image** retrieval; content based **image** retrieval; **image** **database** retrieval; vector distance measures; color segmentation; color vectors extraction; **image** indices; similarity; query color; directional measures; human color perception

Class Codes: B6135 (Optical, image and video signal processing); C7250R (Information retrieval techniques); C6160S (Spatial and pictorial databases ); C5260B (Computer vision and image processing techniques)

Copyright 1999, IEE

10/5/21 (Item 9 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

5520040 INSPEC Abstract Number: C9704-6160S-014

**Title: Fast retrieval of cursive handwriting**

Author(s): Kamel, I.

Author Affiliation: Matsushita Inf. Technol. Lab., Princeton, NJ, USA

Conference Title: Proceedings of the 1996 ACM CIKM. International Conference on Information and Knowledge Management p.91-8

Editor(s): Barker, K.; Ozsu, M.T.

Publisher: ACM, New York, NY, USA

Publication Date: 1996 Country of Publication: USA xi+344 pp.

ISBN: 0 89791 873 8 Material Identity Number: XX96-02903

U.S. Copyright Clearance Center Code: 0 89791 873 8/96/11.\$3.50

Conference Title: Proceedings of 5th International Conference on Information and Knowledge Management

Conference Sponsor: ACM; NASA; Bell Commun.; NSF; AAAI; IEEE Comput. Soc. ; et al

Conference Date: 12-16 Nov. 1996 Conference Location: Rockville, MD, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: Presents an indexing method that can be used to search a large collection of cursive handwriting. The basic idea is to segment each cursive string into a set of strokes. Each of these strokes can be described with a set of features and can thus be stored as points in the feature space. The Karhunen-Loeve transform is then used to minimize the number of features used (the data dimensionality), and thus the **index** size. Feature **vectors** are stored in an R-tree. **Similarity searching** can be performed by executing a few range queries and then applying a simple voting algorithm to the output in order to select the strings that are most similar to the **query** . The proposed index can support **similarity queries** as well as substring matching. It is resilient to the kind of errors that result from the segmentation process, namely stroke insertion/deletion and m-n substitution. The proposed index achieves a substantial saving in search time over a sequential search. Moreover, it improves the matching rate by up to 46% over the sequential search. (11 .Refs)

Subfile: C

Descriptors: handwriting recognition; **image** segmentation; indexing; query processing; software performance evaluation; string matching; transforms; tree searching; visual **databases**

Identifiers: fast retrieval; cursive handwriting; indexing method; cursive string segmentation; stroke set; feature space; Karhunen-Loeve transform; feature number minimization; data dimensionality minimization; index size minimization; feature vectors; R-tree; **similarity searching** ; range queries; voting algorithm; **similarity queries** ; substring matching ; error resilience; stroke insertion; stroke deletion; m-n substitution; search time; matching rate

Class Codes: C6160S (Spatial and pictorial databases); C6120 (File



10/5/22 (Item 10 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2003 Institution of Electrical Engineers. All rts. reserv.

4725434 INSPEC Abstract Number: C9409-6160S-018

**Title: A query based mechanism for multimedia information retrieval**

Author(s): Davcev, D.; Cakmakov, D.; Arnautovski, V.

Author Affiliation: Fac. of Electr. Eng. & Comput. Sci., Skopje Univ.,  
Macedonia  
p.21-38

Publisher: Arizona State Univ, Tempe, AZ, USA

Publication Date: 1992 Country of Publication: USA 188 pp.

Conference Title: Proceedings of Workshop on Multimedia Information  
Systems

Conference Sponsor: Arizona State Univ.; Syracuse Univ.; Univ. Kentucky

Conference Date: 7 Feb. 1992 Conference Location: Tempe, AZ, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

**Abstract:** The basic elements of a multimedia cognitive-based information retrieval system called AMCIRS which integrates **image** and text information have been described by D. Davcev et al. (1991). We extended the AMCIRS query based mechanism which is based on multimedia objects content search using the vector model. The content **search** process is deduced to the **similarity** estimation between **query** and **index vectors**. An important part of this process is the similarity estimation between geometrical **image** forms found in the query and **index vectors**. The main objective of the paper is to introduce the similarity estimation model (SEM) for geometrical objects as a part of a query mechanism of AMCIRS. The model for polygon similarity estimation introduces a numerical measure of similarity between two polygons and gives acceptable results for all polygon forms and any number of vertices. The algorithm based on this model as well as the simulation results are also given. (20 Refs)

Subfile: C

Descriptors: computational geometry; information retrieval systems;  
multimedia systems; query processing

Identifiers: query based mechanism; multimedia information retrieval;  
multimedia cognitive-based information retrieval system; AMCIRS query based  
mechanism; text information; multimedia objects content search; vector  
model; content search process; similarity estimation; **index vectors** ;  
geometrical **image** forms; SEM; geometrical objects; polygon similarity  
estimation

Class Codes: C6160S (Spatial and pictorial databases); C7250 (  
Information storage and retrieval); C4260 (Computational geometry)

10/5/23 (Item 1 from file: 94)

DIALOG(R)File 94:JICST-Eplus

(c)2003 Japan Science and Tech Corp(JST). All rts. reserv.

04845779 JICST ACCESSION NUMBER: 01A0297623 FILE SEGMENT: JICST-E  
**High-dimensional Indexing Methods for Similarity - search of Image  
Contents.**

KAMIKAWA NOBUHIKO (1); IWASAKI KAZUMASA (1)

(1) Hitachi, Ltd.

Joho Shori Gakkai Ronbunshi (Transactions of Information Processing Society  
of Japan), 2001, VOL.42,NO.SIG1(TOD8), PAGE.140-147, FIG.12, REF.5

JOURNAL NUMBER: Z0778AAZ ISSN NO: 0387-5806

UNIVERSAL DECIMAL CLASSIFICATION: 681.3:061.68 681.3:621.397.3  
002.5:005

LANGUAGE: Japanese COUNTRY OF PUBLICATION: Japan

DOCUMENT TYPE: Journal

ARTICLE TYPE: Original paper

MEDIA TYPE: Printed Publication

**ABSTRACT:** Recently, requirement of fast **similarity - search for images**

is increasing. The **indexing** method for n-dimensional **vector** has great interest, because methods of the **similarity - search** for **images** usually use n-dimensional vectors have features of the **image** to calculate similarity. A feature vector of an **image** often has dimensions over a hundred but precedent method of the **similarity - search** are not work effectively for High-dimensional vectors. In this paper we propose a method of the **similarity - search** who can work effectively for High-dimensional vectors. (author abst.)

DESCRIPTORS: **image** retrieval; **database** ; tree structure; similarity; vector space; response time; accuracy; tree(graph)  
IDENTIFIERS: similar **image** retrieval; B tree  
BROADER DESCRIPTORS: retrieval; structure; property; mathematical space; space; time; degree; subgraph; graph  
CLASSIFICATION CODE(S): JD03030U; JE04010I; AC06020S

10/5/24 (Item 2 from file: 94)

DIALOG(R) File 94:JICST-EPlus

(c)2003 Japan Science and Tech Corp(JST). All rts. reserv.

04845768 JICST ACCESSION NUMBER: 01A0297612 FILE SEGMENT: JICST-E  
Visual Database . Texture Image Retrieval Based on the Hierarchical  
Correlations of Wavelet Coefficients.

KOBAYAKAWA MICHIMIRO (1); HOSHI MAMORU (1); OMORI TADASHI (1)

(1) Dentsudai Daigakuin Johoshisutemugakukenkyuka

Joho Shori Gakkai Ronbunshi (Transactions of Information Processing Society  
of Japan), 2001, VOL.42, NO.SIG1(TOD8), PAGE.12-20, FIG.2, TBL.3, REF.11

JOURNAL NUMBER: Z0778AAZ ISSN NO: 0387-5806

UNIVERSAL DECIMAL CLASSIFICATION: 681.3:061.68 681.3:621.397.3  
002.5:005

LANGUAGE: Japanese COUNTRY OF PUBLICATION: Japan

DOCUMENT TYPE: Journal

ARTICLE TYPE: Original paper

MEDIA TYPE: Printed Publication

ABSTRACT: In this paper we propose a robust texture **image** retrieval using hierarchical relations between the decomposed subimages obtained by wavelet transform. Key idea is to describe texture information in terms of the hierarchical correlations between the wavelet coefficients of the adjacent level. Firstly, we express the pyramidal structure of wavelet coefficients by associating the nodes of a complete quad tree with the wavelet coefficients. Secondly, we define a hierarchical dissimilarity vector between a parent node and his child, to express a hierarchical relation between them. Thirdly, to describe a relation among child nodes, we compute a covariance matrix of the dissimilarity vectors. We associate the covariance matrix with the parent node. We define the texture vector by the diagonal of element of the covariance matrix. And then define the texture feature vector of level 1 by the pair of the mean and the standard deviation of texture vectors of level 1. Finally, by applying the discriminant analysis to the set of the texture feature **vectors** , we make an effective **index** of the **database** . For retrieving similar **images** , we use the k-nearest neighbor **search** in the index space. The **similarity** between two **images** is defined by the Euclidean distance between the corresponding feature vectors of the **images** . To evaluate the performance of the retrieval, we made experiments on "Cloth Collections" consisting of 51 textile patterns with 10 different resolutions ( **image** size is 1024\*1024 pixels). The experiments showed that the performance of retrieval is good and that the proposed method is robust with respect to resolution. (author abst.)

DESCRIPTORS: **image** **database** ; **image** retrieval; wavelet transform; information retrieval system; texture processing; quadtree; interactive processing; similarity; resolving power; performance evaluation

IDENTIFIERS: similar **image** retrieval

BROADER DESCRIPTORS: **database** ; retrieval; mathematical transformation; mapping(mathematics); transformation and conversion; information system ; computer application system; system; treatment; tree(graph); subgraph ; graph; property; performance; evaluation

CLASSIFICATION CODE(S): JD03030U; JE04010I; AC06020S

10/5/25 (Item 3 from file: 94)

DIALOG(R) File 94:JICST-EPlus

(c)2003 Japan Science and Tech Corp(JST). All rts. reserv.

04718315 JICST ACCESSION NUMBER: 00A1003969 FILE SEGMENT: JICST-E

Similarity Search Index Using Vector Approximation VA-TREE.

YOSHIDA TADASHIRO (1); AKAMA HIROKI (1); TANIGUCHI NOBURO (1); YAMAMURO

MASASHI (1); KUSHIMA KAZUHIKO (1)

(1) Nippon Telegraph and Telephone Corp. (NTT), Cyber Space Lab., JPN  
Joho Shori Gakkai Ronbunshi(Transactions of Information Processing Society  
of Japan), 2000, VOL.41,NO.SIG6(TOD7), PAGE.1-11, FIG.9, TBL.3, REF.20

JOURNAL NUMBER: Z0778AAZ ISSN NO: 0387-5806

UNIVERSAL DECIMAL CLASSIFICATION: 681.3:061.68 002.5:005

LANGUAGE: Japanese COUNTRY OF PUBLICATION: Japan

DOCUMENT TYPE: Journal

ARTICLE TYPE: Original paper

MEDIA TYPE: Printed Publication

ABSTRACT: This paper describes the **similarity search index** VA-TREE,  
which uses the **vector** approximation method to represent internal and  
leaf nodes in a tree structure index. Our experiment on 16-dimensional  
hue, 16-dimensional intensity and 24-dimensional shape data used in an  
actual **image** retrieval application shows the advantage over the flat  
structure **index** VA-File, which uses **vector** approximation as well.  
The experiment also shows that the number of accessing feature vectors  
on a disk in VA-TREE is not affected by data distribution as much as  
VA-File or VAM Split R-tree. (author abst.)

DESCRIPTORS: **database** ; data compression; tree structure; data retrieval;  
tree(graph); index; multi-media; **image database** ; **image** retrieval;  
hue; luminance

IDENTIFIERS: multimedia **database**

BROADER DESCRIPTORS: data processing; information processing; treatment;  
structure; fact retrieval; information retrieval; retrieval; subgraph;  
graph; information media; photometric quantity

CLASSIFICATION CODE(S): JD03030U; AC06020S

10/5/26 (Item 4 from file: 94)

DIALOG(R) File 94:JICST-EPlus

(c)2003 Japan Science and Tech Corp(JST). All rts. reserv.

03041521 JICST ACCESSION NUMBER: 97A0019593 FILE SEGMENT: JICST-E

**A Method for Indexing Feature Vectors and Its Related Searching  
Approach.**

YAO Z (1); HAMADA TAKASHI (2)

(1) SECOM Information System; (2) Nat. Center for Sci. Inf. Syst.

Joho Shori Gakkai Ronbunshi(Transactions of Information Processing Society  
of Japan), 1996, VOL.37,NO.11, PAGE.2056-2064, FIG.7, REF.18

JOURNAL NUMBER: Z0778AAZ ISSN NO: 0387-5806

UNIVERSAL DECIMAL CLASSIFICATION: 681.3:061.68

LANGUAGE: Japanese COUNTRY OF PUBLICATION: Japan

DOCUMENT TYPE: Journal

ARTICLE TYPE: Original paper

MEDIA TYPE: Printed Publication

ABSTRACT: For the similarity retrieval based on feature **vectors**, how to  
construct an **index** of feature **vectors** to improve the retrieval  
efficiency has become an important topic. In this article we introduce  
a newly developed **index** method of feature **vectors**, which is  
composed of "recurrence clustering" and "removal search strategy."  
Recurrence clustering is a classification method used to construct a  
tree-like **index** of feature **vectors** based on **similarities** between  
feature vectors, and removal **search** strategy is a method  
tailor-developed to suit for searching the index structure constructed  
by recurrence clustering. The similarity retrieval based on feature  
vectors can be efficiently improved by this new approach. That is, the  
retrieval cost of this approach is less than that of linear associative  
retrieval strategy when the recall ratio is less than 10%. The

effectiveness of the approach was confirmed in relative experiments  
using feature vectors extracted from full-text. (author abst.)

DESCRIPTORS: **database** ; information retrieval; similarity;  
vector(mathematics); clustering; search problem; tree search; tree  
structure; statistical method; full-text search

BROADER DESCRIPTORS: retrieval; property; linear algebra; algebraic system;  
modification; problem; structure; document retrieval

CLASSIFICATION CODE(S): JD03030U

?